"Now that you mention it, I wonder..." : Awareness, attention, assumption

de Jager, S.T.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 5
Pragmatics of decision-making

The question of her decision is one not to be lightly considered, and it is not for me to presume to set myself up as the one person able to answer it. And so I leave it with all of you: Which came out of the opened door — the lady, or the tiger?

Frank Stockton, “The Lady or the Tiger?”

We began this dissertation with a story about Walt’s difficult Saturday morning. Late for a job interview and frantically searching for his car keys, he was helped out by Perky Pat:

(1) Did you leave them in the car when you came in drunk last night?

Walt’s immediate reaction was to run to the car and check, and by now we have all the tools we need to understand and model his behaviour. Still something is missing: we take Pat’s utterance as constituting advice for Walt to search the car, but we can only give an ad hoc and intuitive description of why this is so. In a similar fashion, “If you take the bicycle you might get a flat tyre” strikes us intuitively as advice against taking the bicycle, but we would be hard pressed to give an account of why without appealing to the very intuitions we are trying to justify.

A more famous example comes from Grice [Gri67, p. 32]:

(2) a. A: I am out of petrol.
   b. B: There is a garage around the corner.

Informally it is easy to see what speaker B intends to convey; Grice derives semi-formally the information that the garage is open via his maxims of cooperative conversation. However a fully formalised account is surprisingly difficult to achieve (in part because of the vagueness of Grice’s formulations; as we will see below, trying to formalise reasoning by relevance immediately raises the question of what formal property relevance consists in).

In recent years game theory and decision theory have been fruitfully applied to such problems (see for example the collection [BJRos5]), as a formal means of modelling the agent-based nature of discourse (the background context
distinguishing a sentence ‘in the abstract’ from an utterance made by an agent and to a purpose, which makes pragmatic reasoning possible).

Game theory is perhaps the more popular approach of the two, as it is a particularly good fit for the complex multi-agent epistemic reasoning involved in linguistic coordination, whether conventional [Lew69] or ‘online’ (e.g., particularised conversational implicature [BR07]). However the very wealth of game-theoretic possibility can be problematic. In coordination games, for instance, standard game-theoretic techniques predict linguistic anti-coordination (in which the speaker makes meaningless noises at random and the hearer ignores him) as a possible ‘convention’. Much effort goes into excluding such absurd cases, before we get to the more interesting pragmatic issues. For certain phenomena in pragmatics —notably those involving coordination problems, whether implicit or explicit— we cannot get by without the rich interactional and epistemic space provided by game theory. Before turning to game theory, however, it is interesting to ask whether the questions we are asking can be answered in a model without the rich multi-agent structure that makes game theory both powerful and difficult to work with.¹

And indeed in our case we can. We will apply the framework of DECISION-THEORETIC PRAGMATICS, which uses the single-agent decision-making model of statistical decision theory, embedded in a multi-agent discourse context that remains mainly notional. The model of decision-making is rich enough to allow complex reasoning about speaker motivations (“Why did she tell me this, and not that?”) of the kind driving the majority of accounts of pragmatic reasoning explicitly based on Grice’s work.

It does not, however, explicitly describe unawareness. The representations of previous chapters were all motivated by the intuition that Pat’s utterance is intended to make Walt aware of a possibility he is overlooking. The fact that it constitutes advice cannot be separated from this intention, so we will have to find a way to combine decision-theoretic pragmatics with an unawareness model.

As I have argued throughout this dissertation, the best way to achieve such a combination is by the design of ‘modular’ theories that can be combined without significant alteration. The model I present here falls somewhat short of that ideal; the structures necessary to represent unawareness are significantly more complicated than those usually employed in decision-theoretic pragmatics, and not just in terms of the components strictly necessary for the representation of awareness itself (the atomic formulae attended to, the ordering generating assumptions, and so on). As was the case for counterfactuals (where the ordering structure required by the awareness account gave support for a causal

¹For a cutting-edge example of both the difficulties and the opportunities provided by rich game-theoretic models, see [Fra09].
analysis of counterfactual similarity) I take this to be no disadvantage for the theory; rather it tells us something more about the right way to think of the simpler structures in common use. Indeed, the parallel between the two cases turns out to be quite close in specific terms, as well as at this level of generality; the counterfactual semantics I gave in Chapter 4 is formally very close to the account of action and agency that underpins the decision-theoretic model.

This chapter grew out of joint work with Michael Franke, which will only see publication after this dissertation is printed and defended [FJ08]. I have revised the formalism substantially, in part to bring it more in line with the systems of the previous chapters and in part because my own views have changed in the interim. I must however record my conceptual debt to Michael; in particular, I owe entirely to him the notion that standard decision theory is riddled through with assumptions about unawareness, which constrain the theory in artificial ways because they remain implicit. It is with this idea that we will begin.

1 Decision theory: a theory in need of unawareness

Bayesian decision theory is a normative theory of decision-making under uncertainty. The decision-maker holds probabilistic beliefs (a probability distribution over possible states of the world) and must choose from a set of possible actions; a particular state of the world paired with a particular action is associated with a numerical utility expressing relative desirability (thought of as the desirability of performing that action if the world is in that state). The normative theory concerns how these elements should be combined to calculate the best choice of action (I will give more formal definitions in a moment).

A decision problem is a particular instantiation of this pattern, representing a particular choice the decision-maker has to make. For instance, we might model the decisions of an investor choosing which projects and companies to fund, or more prosaically, the choice of how to travel to work in the morning. While awareness is not explicitly a part of a traditional decision problem, it is implicitly absolutely pervasive.

The states of the world that are chosen to represent a particular problem are picked out by attending to some distinctions and ignoring others (the finegrainedness aspect of awareness). The weather is a relevant variable for choosing how to get to work, but completely irrelevant for investment decisions; the modeller must choose which distinctions to attend to and which to ignore. As well as finegrainedness, traditional decision problems embody assumptions: not every conceivable possibility makes it into the states under consideration. Many of these assumptions will be entirely innocent; when choosing how to get to work it is probably safe to assume that there will not be a sudden and devastating flood, or similar madness. However this need not always be the case. An investor may well consider the possibility that each company
she is considering funding might fail, but nonetheless be unprepared for a widespread financial collapse such as the recent credit crisis.

Similar remarks apply to the set of available actions. The set is restricted by something analogous to assumption (there is no point including actions such as “Invest in plastics R&D” if the problem at hand is getting to work on time; less innocently, the agent might be unaware of relevant possible actions, for instance if she does not realise that a new bus line has started operating), and differentiated at a level of finegrainedness that reflects the problem at hand (cycling might count as one action, or be split into the options “take the racing bike” and “take the slow-and-sturdy city bike”).

Designing a decision problem, then, is really a kind of attention-modelling task. The modeller must try to choose the right actions and states (drawing attention to the right possibilities) so that the decision problem represents the relevant alternatives while ignoring irrelevant possibilities and distinctions.

Once one takes this perspective, however, an obvious question arises: how are we to represent changes in the awareness of a decision-maker, as she comes to recognise new possibilities for action or new ways the world might be? These changes in awareness are an important part of real human decision-making, after all; sometimes the most helpful thing you can do for someone who is struggling with a difficult decision is simply to point out an option they haven’t thought of. Because the representation of unawareness in a particular decision problem is only implicit, this kind of change is impossible to model systematically without extending the theory.

In the first part of this chapter I will extend one particular model of decision theory with explicit unawareness, so that changes in the awareness state of the agent can lead to evolution of the decision problem she faces. The aim for this section is to enrich standard decision-theoretic representations so that our agents can engage in problem-solving dialogues like those in (1) and (2), and so that changes in the agent’s awareness of possibilities systematically carry over into her decision problem.

In Section 4 I will put these enriched models to work on some linguistic issues. Decision-theoretic pragmatics uses the normative theory of decision-making to model pragmatic reasoning about cooperative discourse. By replacing classical decision theory with the awareness-enriched version we can extend the range of decision-theoretic pragmatics without making any changes to the core theory itself. Finally in Section 6 I sum up and mention a particularly interesting loose end.

2 · Models for decision-making

We will start with the simple representation of a decision problem which is generally accepted in decision-theoretic pragmatics. (In the wider world of Bayesian or statistical decision theory various alternatives exist, most of which
I will simply ignore for the sake of simplicity.) We need first some basic (and entirely standard) notation for probabilities.

**Definition 5.1**: Probability basics. Let $S$ be a countable set. A **probability distribution over** $S$ **is a function** $P$ **from** $S$ **to the real interval** $[0, 1]$ **such that** \( \sum_{s \in S} P(s) = 1 \). We write \( \Delta(S) \) for the set of all such distributions. A distribution on $S$ induces a probability measure on the \( \sigma \)-algebra of subsets of $S$ (also known as events), which we also notate with $P$: if $X \subseteq S$ is any subset of $S$, then

$$P(X) = \sum_{s \in X} P(s).$$

If $P$ is a probability distribution over $S$ and $X, Y$ are events (subsets of $S$) then $P(X | Y)$ represents the conditional probability of $X$ given $Y$, given by

$$\frac{P(X \cap Y)}{P(Y)}$$

(and of course only well-defined when $P(Y) \neq 0$).

A decision problem formally incorporates the elements given informally above: states of the world and probabilistic beliefs about them, possible actions, and a utility function representing desirability.

**Definition 5.2**: Decision problem. A **decision problem** is a structure with four components, $D = \langle S, A, P, U \rangle$ where

- $S$ is a finite set of states (not to be confused with possible worlds; states are typically partial objects, recording only a few features relevant for the problem at hand);
- $P$ is a probability distribution over $S$;
- $A$ is a finite set of possible actions;\(^3\) and
- $U : S \times A \to \mathbb{R}$ is a utility function: $U(s, a)$ gives the numerical desirability of taking action $a$ if the state of the world is $s$.

We will call a state $s \in S$ proper if $P(s) > 0$.

\(^2\)Economists are often interested in continuous action spaces representing price offers, production quantities, or other notionally real-valued quantities. The applicability of unawareness to such actions is unclear, since different possible values are not separate concepts in the same way that, say, “cycle to work” and “take the tram” are. Taking a discrete (and finite) action space is also convenient in that it allows me to reuse the awareness machinery developed in the previous chapters without substantial revision.

\(^3\)Note that this is not the assumption set, represented by $A$ in previous chapters. I will adjust the notation for assumptions slightly so as not to mix ‘$A$’ and ‘$\mathcal{A}$’, in an attempt —nonetheless probably futile— to avoid confusion.
Given such a problem we can calculate the set of best actions to take by calculating the expected utility of each action (the average utility the action will earn, if the states are distributed according to \( P \)). The following definitions are also standard, except for average expected utility (about which more below).

**Definition 5.3: Expected utility and related notions.** Let \( D = (S, A, P, U) \) be a decision problem, and \( a \in A \) an action. The expected utility of \( a \) in \( D \) is given by

\[
EU_D(a) = \sum_{s \in S} (P(s) \cdot U(s,a)).
\]

The average expected utility of a set of actions is the average of their individual expected utilities (equivalent to expected utility under uniform selection of the action): let \( X \subseteq A \), then

\[
AEU_D(X) = \frac{\sum_{a \in X} EU_D(a)}{|X|}.
\]

The best actions in \( D \) are the actions that maximise expected utility:

\[
BA(D) = \{ a \in A : \forall a' \in A : EU_D(a) \geq EU_D(a') \}.
\]

The value of \( D \) is the average expected utility of its set of best actions:

\[
val(D) = \frac{AEU_D(BA(D))}{|BA(D)|}.
\]

The average expected utility of a set of actions demands some comment. It corresponds to a random decision by the agent according to a uniform probability distribution over the set of actions; a biased probability distribution would give a different result, whenever the expected utilities of the actions in the set vary. For calculating the value of a decision problem this is unimportant since all the actions under consideration (the best actions of the decision problem) have the same expected utility (and thus the choice of distribution makes no difference, and the uniform distribution might as well stand in for the others). We will see cases later on, however, where this assumption plays a real role.

Thus far we have decision theory with only implicit unawareness. The tactic I want to apply is to define a richer model (incorporating awareness but also with some additional structure) from which a decision problem can be ‘read off’.

### 2.1 Richer decision models

The eventual aim of this chapter is to be able to model the effect on a problem-solving conversation produced by changes in awareness. Typical conversational moves would be utterances like:

(3) a. Did you think of taking the bicycle?
b. There might be a tram strike.

c. If you take the bicycle you might get a flat tyre.

d. The forecast is for rain; if you bicycle you will get wet.

I will give a model that is rich enough to support the beginnings of a compositional analysis of such updates, although I will not attempt such an analysis in detail. The model is rather complex, so we will construct it in two stages.

In the first stage we define the structures necessary to talk about the state of the world. “If you take the bicycle you might get a flat tyre” talks explicitly about a possible outcome of an action, which is nowhere explicitly represented in a decision problem in the format given above. This is no defect for standard decision theory, since the utility is all that is needed for the calculation. However we want our agents to be able to discuss outcomes, so they need to be present somewhere in the model. (I will also talk about ‘worlds’ rather than ‘states’, for reasons which will become clear in Section 2.2.)

**Definition 5.4:** Enriched model part 1: states, probabilities, utilities. An enriched model is a structure $E = \langle W, \Omega, V, P, \ldots, U \rangle$ containing the following components:

- $W$ is a finite set of worlds, $\Omega$ is as always a set of atomic formulae, and $V$ is a valuation function on worlds: $V : W \times \Omega \rightarrow \{0, 1\}$.
- $P$ is a probability distribution over $W$, representing the agent’s epistemic uncertainty about the current state of the world.4
- $U : W \rightarrow \mathbb{R}$ is a bounded utility function on worlds.

So far we have enough material for the agent to express her information and uncertainty (via the probability distribution), and her preferences (via the utility function). She is not yet, however, agentive in the sense of being able to take action in the world of the model.

**Definition 5.5:** Enriched model part 2: actions. An enriched model also contains two more components: $E = \langle W, \Omega, V, P, A, O, U \rangle$ is a fully specified enriched model, where $A \subseteq \Omega$ is a set of special atomic formulae known as actions, and $O$ is a set of outcome distributions, one for each world/action pair. For each world $w \in W$ and action $a \in A$, the outcome distribution $O^w_a \in \Delta(W)$ is a probability distribution representing the possible results of taking action $a$ in the world $w$; if $O^w_a(w') = 0.3$, for instance, then taking action $a$ in $w$ leads to the world $w'$ with probability 0.3. The only constraint on the distribution is that if $O^w_a(w') > 0$, then $V(w', a) = 1$ (one guaranteed outcome of taking an action is that the action is taken).

4I will not give any formal representation of temporal structure, however it is important to bear in mind that elements of $W$ may represent either current states of the world or states that the world may be brought to through the actions of the agent. The probability distribution represents current information; it may be that a world that is assigned zero probability mass is nonetheless reachable via an action.
The outcome distributions reflect the notion of \textit{metaphysical uncertainty} that we saw in Chapter 4. Some actions (such as rolling a die) are inherently indeterminate, and the most we can say about their results is that they will be distributed according to a particular statistical pattern. Of course \textit{epistemic} uncertainty about the results of actions also has a place in the model: if the die may or may not be loaded, then the outcome distributions differ for the world where it is fair and the world where it is not. We will return to this notion below.

Besides metaphysical uncertainty, the main idea of the definition is that actions \textit{change the world}. Trivially, taking the bicycle to work changes the world into one in which the agent took the bicycle to work; typically various other elements of the world will change along with this one (she may get windblown hair and tired legs, and the bicycle will end up locked in the carpark rather than at home in the cellar, for instance). If \( h \) is the propositional formula saying that the agent has windblown hair, and \( [h] \subseteq W \) is the set of worlds satisfying that formula (the event, or proposition), then it is quite reasonable that in her decision problem \( P([h]) = 0 \). This means that her information tells her that \textit{at this moment} she does not have windblown hair; similarly, at this moment she is not bicycling to work (the action \( b \)). The conditional probability \( P([h] \mid [b]) \) is undefined (since \( P([b]) = 0 \)) and anyway does not represent the probability that bicycling leads to windblown hair; it represents the probability that the agent has windblown hair \textit{now}, conditional on her \textit{now} being on a bicycle.

The probability that bicycling to work leads to windblown hair is read off instead by first \textit{changing} the model with action \( b \), then looking at the probability of \( h \) in the resulting distribution. The notion was called \textit{general imaging} in [Gär82],\textsuperscript{5} and is defined (in our setting) as follows.

\begin{definition}
Enriched model: results of action. \textit{Take an enriched model} \( E = \langle W, \Omega, V, P, A, O, U \rangle \), \textit{and any action} \( a \in A \). \textit{Then} \( E[a] \), \textit{the result of taking action} \( a \) \textit{in} \( E \), \textit{is a new enriched model:}

\[ E[a] = _d \langle W, \Omega, V, P[a], A, O, U \rangle \]

\textit{in which the probability distribution over worlds is given by:}

\[ P[a](w') = _d \sum_{w \in W} (P(w) \cdot O^w_a(w')) \]

This updated probability represents the chance of \textit{arriving} at \( w' \) by performing action \( a \), \textit{while the original} \( P \) \textit{represented the chance that the world is already} \( w \).
\end{definition}

\textsuperscript{5}Lewis defined imaging in [Lew76], however he was applying it to Stalnaker’s counterfactual semantics, with its assumption (validating conditional excluded middle) that each world has only a single closest \( \varphi \)-neighbour. The generalisation given by Gärdenfors allows sets of \( \varphi \)-neighbours, in our case representing metaphysical uncertainty about the outcome of an action.
It is this updated probability we will use in calculating the expected utility of an action.

Seen in this light, the outcome distributions are closely related to a counterfactual similarity relation, especially when the latter is generated by a causal semantics as I have argued in the previous chapter. The counterfactual closest \( \varphi \)-worlds to \( w \) incorporate the minimal expected causal consequences if \( w \) had been adjusted so that \( \varphi \) was the case; the decision-theoretic \( b \)-outcome worlds of \( w \) incorporate the minimal expected causal consequences if \( w \) is adjusted by making \( b \) the case. Outcome distributions are simply a probabilistic incarnation of the same idea of causal similarity driving the counterfactual semantics that we have already seen (indeed, imaging was first proposed for a probabilistic treatment of counterfactual beliefs).

There is one significant element left out of the model: temporal structure in the worlds themselves. An element of \( W \) should really be a world-time pair, so that our agents can properly discuss knowledge about the past and future; taking an action does not, in fact, ‘change the world’ in its entirety but only the available future course of events, and so on. I will deal with these complications mainly informally; the only temporal dimension explicitly represented in the model is the distinction between the ‘now’ of epistemic uncertainty and the ‘future’ of results of actions.

With that proviso, let us see the connection between the enriched model and the decision problem as standardly conceived.

**Definition 5.7: Impoverishing an enriched model.** Let \( E \) be an enriched decision problem with \( E = \langle W, \Omega, V, P, A, O, U \rangle \). Define the **impoverishment** of \( E \), a standard decision problem \( D_E = \langle S_D, A_D, P_D, U_D \rangle \), as follows:

- \( S_D = W \) (the states are simply the worlds from the enriched model);
- \( A_D = A \) (the same actions are reused);
- \( P_D = P \) (the probability distribution representing epistemic uncertainty is also reused);

---

6Outcome distributions are more direct probabilistic analogues of selection functions: for each world \( w \) and formula \( \varphi \), \( f(\varphi, w) \) gives directly the closest world(s) to \( w \) where \( \varphi \) holds. (Stalnaker introduced the notion on the assumption that a single world was closest, validating the conditional excluded middle; Lewis generalised it to allow a set of equidistant closest \( \varphi \)-worlds, which is the formulation we take here.) Every similarity ordering can be represented by a family of selection functions, but the converse does not hold. An ordering carries additional information about similarity, such as that it is transitive, which must be expressed as a constraint on sets of selection functions. I assume implicitly, for simplicity in discussion, that actions are mutually incompatible, so that there is never any reason to compare the relative similarity of the closest \( a \)-world and the closest \( b \)-world to \( w \) (where \( a \) and \( b \) are actions). Allowing mutually compatible actions would simply require putting the same constraints (expressed probabilistically) on combinations of outcome distributions that similarity orderings impose on selection functions.
Chapter 5 · Pragmatics of decision-making

• $U_D$ is given by, for each $w \in S_D$ and $a \in A_D$:

$$U_D(w, a) = \sum_{w' \in W} (O^a_w(w') \cdot U(w')).$$

The impoverished decision problem is clearly very closely related to the enriched model it derives from. The only real difference between the two representations is that the impoverished model hides the detailed structure of (expected) outcomes behind the numerical calculation of utility. However, this kind of impoverishment, with all its fine detail, is not generally what we will want. An enriched model contains full possible worlds rather than partial states, so it will typically be much more fine-grained than our agent’s mental state. We need a way to turn the total structures of the enriched model into partial structures, keeping only the distinctions (and actions) that the agent is aware of.

2.2 · Impoverishment via unawareness

Just as for the dynamic model of Part I, the strategy will be to treat the enriched model as a ‘model of reality’, and restrict the agent’s view of this model according to her state of (un)awareness. Unlike in Part I, however, the emphasis here is on partiality rather than assumption. Typically the states of the impoverished decision problem will aggregate a large number of worlds, erasing the distinctions between them in the awareness-limited view of the agent. And just as I argued in Chapter 1 for unawareness of objects, the ‘gap model’ (with unawareness corresponding to an ‘assumption of absence’) is particularly appropriate for the actions of a decision problem. If the agent is unaware of the possible action “Take bus 405” (whether because she has not thought of buses at all, or because she does not know of that particular line), that action should simply be absent from the impoverished model representing her view of the decision problem she faces.

The two modes of unawareness (assumption and finegrainedness) correspond to two kinds of filters that we will apply to get from an enriched model to an impoverished decision problem under unawareness: assumptions provide restrictions, and finegrainedness provides aggregation.

**Definition 5.8**: Awareness state. Fix an enriched model $E = (W, \Omega, V, P, A, O, U)$. Similar to the definition of Chapter 2, an awareness state is a pair $\sigma = (W_\sigma, \Xi)$ where

- $W_\sigma \subseteq W$ represents the assumptions (relabelled so as to avoid both ‘A’ and ‘A’ as components of the similar structures);

- $\Xi \subseteq \Omega$ is as before the set of proposition letters the agent attends to.

Recall that $A$, the set of actions, is nothing more than a specially-treated subset of $\Omega$. Likewise, we can single out the actions the agent is aware of: $A_\sigma = A \cap \Xi$. 

110
Unlike the awareness states of the previous chapters, we don’t need to include beliefs. These are encoded in the prior probability distribution over the enriched model as a whole, and we will derive probabilistic beliefs under assumptions simply by conditionalising on the assumption set $W_\sigma$. The idea is that the distribution over the enriched model records the probabilistic beliefs the agent would have, if she paid proper attention to all possibilities — a sort of limit case of total attention. To make this precise, we need to see how to read off an impoverished decision problem via an awareness state.

**Definition 5.9: Impoverishment via unawareness.** Let $E$ be an enriched model with $E = \langle W, \Omega, V, P, A, O, U \rangle$ and $\sigma = \langle W_\sigma, \Xi \rangle$ an awareness state. Then the impoverishment of $E$ under $\sigma$ is a decision problem

$$D_E^\sigma = \langle S_D, A_D, P_D, U_D \rangle.$$ We need a subsidiary notion to define the states. Let $w_1, w_2 \in W$ be two worlds. Then $w_1$ and $w_2$ are **equivalent under** $\Xi$, written $w_1 \equiv_\Xi w_2$, if for each $p \in \Xi$, $V(w_1, p) = V(w_2, p)$. Now the components are given by:

- $S_D = W_\sigma/\equiv_\Xi$; that is, the states are the equivalence classes under $\Xi$ of the worlds the agent entertains;
- $A_D = A_\sigma$ (the actions of the decision problem are just those the agent is aware of);
- $P_D$ is given for each $s \in S_D$ by
  $$P_D(s) = \sum_{w \in s} P(w | W_\sigma) = P(s | W_\sigma)$$
  (that is, the probability distribution treats states as events in the enriched model, conditionalised on the agent’s assumptions); and
- $U_D$ is given for each $s \in S_D$ and $a \in A_D$ by
  $$U_D(s, a) = \sum_{w \in s} \sum_{w' \in W_\sigma} (P(w | s) \cdot O^w_\sigma(w' | W_\sigma) \cdot U(w')).$$

The utility calculation contains nothing unexpected. To find the expected utility of action $a$ in state $s$ we need to take expectations over worlds in $s$ (epistemic uncertainty) but also over outcomes of the action in those worlds (metaphysical uncertainty), before measuring the utility in those outcomes. The most important point is the conditionalisation: the agent cannot see outside $W_\sigma$, so the outcome probabilities need to be scaled accordingly (an agent unaware of the double-zero in American roulette would assign the outcome probability $\frac{1}{37}$ to each of the other numbers zero through 36 on the wheel).
Chapter 5 · Pragmatics of decision-making

The impoverishment transformation encodes the two modes of unawareness: the decision problem includes only worlds from the set $W_\sigma$ (assumption), and the states are distinguished only according to propositions in $\Xi$ (fine-grainedness). Whenever a state contains worlds differing on the value of some proposition letter $p$, we can see the state as a partial object unspecified for the value of $p$. However if the agent later becomes aware of $p$, the partial object that is the state splits into two more completely specified objects, differing in their valuation for $p$.

The resulting decision problem also has the intuitive property that the agent can describe the differences between the different states, using only her language $L^\Xi$. However, just as for awareness states in the previous chapters, it may be malformed in more subtle ways.

**Definition 5.10: Awareness-consistency.** Let $E = \langle W, \Omega, V, P, A, O, U \rangle$ be an enriched model and $\sigma = \langle W_\sigma, \Xi \rangle$ an awareness state.

Just as for our original models, a state is **awareness-consistent** if every valuation of $\Xi$ that occurs in the model also occurs within the assumption set: for all $\varphi \in L^\Xi$,

$$\text{if for some } w \in W, w \models \varphi, \text{ then for some } w' \in W_\sigma, w' \models \varphi.$$ 

The state is **decision-awareness-consistent** if

$$\exists w \in W_\sigma : P(w) > 0$$

*(the assumptions permit probabilistic beliefs; equivalently, $P(W_\sigma) > 0$)*, and

$$\forall w \in W_\sigma \forall a \in A_\sigma O^w_a(W_\sigma) > 0.$$ 

*(results of actions are always probabilistically defined).*

Just as for the simpler models of the previous chapters, we need to define an ordering $\preceq$ on $W$ that provides the assumptions. Not every ordering will produce decision-awareness-consistent impoverished models, because of the two conditions. A natural way to force the assumptions to permit probabilistic beliefs is to require that some elements of the lowest level of the ordering carry probability mass. (Recall that regardless of the awareness state of the agent, the worlds in the lowest level of $\preceq$ are always included in her assumptions.) It is natural that some of these worlds are held possible because of the intuitive correspondence between probabilistic belief and assumption: the lowest level of $\preceq$ represents the way the agent assumes the world is before she gives the matter any thought. If her probabilistic beliefs as encoded ‘in the limit’ of the enriched model should not give these worlds probability mass, this would be at least evidence that her assumption-formation machinery is operating very
On the other hand it is equally natural that some of these lowest-level worlds be held (probabilistically) impossible (and thus a requirement that all the worlds carry probability mass would be too strong): they may represent future situations that are so salient the agent automatically entertains them, despite only holding them possible as the result of actions not yet taken.

More problematic is the requirement that action results are everywhere defined. It might seem that awareness-consistency should guarantee this, since if the agent attends to an action $a$, it provides worlds in which the action $a$ is taken (that is, where $V(w, a) = 1$). The problem is that these need not necessarily be the outcomes of worlds the agent holds possible. For a trivial example, if all the worlds in $W_\sigma$ that satisfy $a$ are (notionally) in the past of $w$, and the agent holds $w$ possible, then from $w$ the action $a$ has no outcome in $W_\sigma$ (since actions cannot ‘time travel’ into the past). I will not attempt to solve this problem with a principled restriction on the structure of $\preceq$ (if such is possible, the expression would seem to be rather complex); instead I will simply require that all impoverished models satisfy the constraints. This will be simple enough to verify by hand, for the examples I am interested in.

## 3 · Updates

The system is to some extent a direct descendant of the models of the first part of the dissertation, so it could in principle be applied to the same kinds of problems. The extra structure of decision theory, however, lets us model advice-giving that talks explicitly about the opportunities for action available to the agent, as in (3).

I will be vague about the specifics of the update mechanism, since it follows the pattern of the simpler system rather closely. For attention updates we replace $W_\sigma$ in the awareness state with an ordering $\preceq$ on $W$, and generate $W_\sigma$ from $\preceq$ and $\Xi$ as in the non-probabilistic model. Belief updates also operate analogously, with non-zero probability corresponding to membership in the belief set $B$ (so that “might $\phi$” holds if some world supporting $\phi$ gets non-zero probability, and the update with a propositional formula $\phi$ sets the probability of each $w_t$ not supporting $\phi$ to zero).

The interesting complication is the action update “If you take the bicycle, you might get a flat tyre.” For this we introduce the connective $\square\rightarrow$, now representing not a counterfactual but an action update (it must be syntactically restricted to take only an action in the antecedent, not a general formula). The reuse of notation is entirely justified, however, since the semantics is exactly analogous to that given in the previous chapter (Definition 4.5, the informational

---

7 This constraint only makes sense when the prior encoded in $E$ represents the ‘start of a conversation’. As the agent accumulates information she will naturally be able to rule out worlds in the lowest level of the association ordering, which we will implement with her updating the prior by conditionalisation.
update version).

That is, an enriched model $E$ supports an ‘action conditional’ $a \rightarrow \varphi$ if $E[a]$ (probabilistically) supports $\varphi$. Updating with $a \rightarrow \varphi$ is only possible if $\varphi$ is propositional (does not contain might) since the update works by eliminating epistemic uncertainty (removing worlds from the set of current epistemic alternatives).\footnote{This is an unfortunate consequence of the ‘double uncertainty’ of these models: a single might can represent either epistemic or metaphysical uncertainty, or even both (recall “If I had taken the fight, I might have won”). The informational update rule that makes propositional updates work gives the wrong results for might conditionals, while the test rule that gives might conditionals the right semantics fails for updates. See the discussion following Definition 4.5 of Chapter 4.} In particular, “If you roll that die you will get a four” is only acceptable if some epistemic alternative contains an unfair die; it cannot be interpreted as pruning outcome distributions (resolving metaphysical uncertainty) but only at the epistemic level.

This is perhaps the simplest probabilistic update system that can be imagined (modulo the awareness dimension and the complexity of the model itself): propositional updates correspond simply to conditionalisation. In particular, although the model can represent subtleties such as “$\varphi$ is more likely than $\psi$”, neither we nor the agents can talk about these systematically (since the object and meta languages are none other than those defined in Chapters 2 and 3).

The awareness updates, however, bring to the fore one final subtlety, which I mention for completeness even though it will play no part in the rest of the chapter: becoming aware of one proposition letter may trigger awareness of another, via a process of association.

3.1 · Associations

Associations were introduced into the formal system of [FJ07] to solve a technical problem. In that paper we distinguished ontologically between actions and proposition letters; becoming aware of an action (such as taking the bicycle) then needed to be associated with becoming aware of at least one potential outcome for that action (having taken the bicycle, as a proposition). This is no longer strictly necessary in the current model, since the action does double duty as a proposition, however it is certainly natural, particularly in cases where an action has two ‘stereotypical’ or ‘expected’ outcomes. Suggesting that the agent flip a coin, for instance, can be reasonably expected to make them aware not just of the action but of the outcomes ‘heads’ and ‘tails’. In fact none of the examples I will consider below have this property, so I will refrain from adding a representation to the formalisation. The notion should probably be added to the simpler models of the previous chapters; a simple way to do so would be to give each proposition letter a set of associations (also proposition letters, trivially including the letter itself), and define the awareness update with $\varphi$ not based on the proposition letters occurring in $\varphi$ but on the union of their
The most important result of this proposal is that we can accommodate the effect of uninformative utterances, in the typical awareness mode. I want to first describe the effect of simply becoming aware of a possibility, without this being induced by hearing a linguistic utterance. The pragmatic effects of uninformative utterances are so important that they tend to overwhelm any intuitions about what other effects they might have. I will try to avoid this effect by framing each example as “Oh! It’s possible that…” or similar; the idea is that the agent has spontaneously become aware of a new possibility, without that possibility being suggested to her by someone else. In Section 4.3 below I show the kinds of pragmatic reasoning this can give rise to.

3.2.1 · Oh! I could take the bicycle!
This awareness update obviously adds an action to $A$ (say, $b$ for “bicycle”), but it must also update the worlds in $W_σ$ accordingly. Becoming aware of $b$ will very likely add some worlds to $W_σ$, and thus change the set of states, but it need not alter the proper states (those that carry probability mass) at all. Remember that these represent what is the case now, while taking the bicycle is an action that (typically) occurs in the future. Typically, then, the worlds added by the awareness update (and the states they aggregate into) will carry zero probability mass in $E$ but positive mass in $E[b]$. The proper states, then, need not change at all, but quite possibly some new worlds will need to be added to $W_σ$, to ensure that every world currently held possible has an outcome (carrying non-zero probability mass after the action update) for $b$. In addition, the action itself becomes available and thus the linkage (via outcome functions) between current states and possible futures becomes richer.

Looking at the impoverished decision problems before and after, the states and probabilities probably will not change; the set of actions grows, and the utilities of the other actions besides $b$ are unlikely to change (only if the states do, in fact), but of course the utility of action $b$ becomes defined at each state.

3.2.2 · Oh! It’s possible that there is a tram strike!
This update has different effects depending on what the agent assumed before she was aware of the possibility of a tram strike. We consider two cases:

---

9I have to hedge with “typically” and so on because it is possible, in principle, to model an agent who becomes aware that she can take a bicycle to work, and simultaneously realises that she is not sure if that is what she is already doing. A famous problem in game theory, the ‘Drunk driver paradox’, concerns a similar scenario (without the awareness dimension) so the notion, though patently absurd, has some application. We will not do any work with such marginal examples, but it is reassuring to know that the system can represent them if pushed to it.
either she was assuming that there would not be a tram strike (a reasonable assumption; the same account holds, *mutatis mutandis*, for the unreasonable assumption that there would be a strike), or she held no such assumption and the finegrainedness effect of unawareness simply left her states unspecified for strikes.

In the first case, the new states she gains will be entirely disjoint from the old ones. The relative probabilities of the existing states will be left unchanged (so if the chance of rain was \( \frac{2}{3} \) on the assumption that there is no strike, the chance of rain conditional on there being no strike is still \( \frac{2}{3} \)). Their absolute probabilities, however, will change as the new states take up probability mass, affecting calculations of best actions and so on.

If the agent did not assume there would be no strike, matters are slightly more complicated. Some or all of her states will have to *split* into halves holding striking and non-striking worlds (and there may still be entirely new states included, for the same reasons as in the simpler case). For a case like a tram strike, splitting states will reveal widely varying utilities across worlds that used to be considered part of one state (taking the tram is of course strongly dispreferred if it doesn’t actually go where you’re going...). My feeling is that for such stark utility differences an assumption is a more realistic modelling choice, because the finegrainedness solution involves averaging over conceptually highly distinct outcomes which the agent nonetheless does not distinguish. On the other hand it is easy to imagine an agent who has not realised that Tram 12 takes two different routes, one of which is longer than the other; her assessment of the utility of “take Tram 12 at 13.10” is naturally seen as an average over her uncertainty about which route that particular service takes, with the averaging kept invisible to her by her unawareness of the distinction.

### 3.2.3 · Oh! If I take the bicycle I might get a flat tyre!

The final example of an awareness update concerns a possibility that can only be realised if the agent takes a particular action.\(^\text{10}\) Mechanically speaking the result is very similar to “There might be a tram strike”: the newly raised possibility adds more worlds, with concomitant changes to the set of states and the possible outcomes. As for adding an action, however, we would typically expect the distribution on (proper) states to remain largely unchanged: the possibility being raised belongs in a future that is only accessible via ‘changing the world’ with the action of taking the bicycle.

So far these cases are relatively straight-forward extensions of update semantics with unawareness into decision theory. We are not yet really modelling problem-
solving discourse, as in (3), though, because these updates completely ignore
the pragmatic force of the utterances. In the next section we take the necessary
further step: the assumed background of a decision problem can provide a
representation of context of utterance, allowing our agents to perform pragmatic
reasoning.

4 · Decision-theoretic pragmatics

Decision-theoretic pragmatics emerges from the observation that conversation
is very often ‘to the point’, where the point in question is some non-linguistic
aim; this is particularly the case in information-gathering scenarios such as
Grice’s famous example given in (2).

In cases like these we might model the first speaker’s situation as a decision
problem (“How to get petrol”) with utilities corresponding to eventual success;
her conversational aim is to increase the value of this decision problem, which
is nothing but a formal way of saying she aims to maximise her chances of
getting her car refueled.

Grice’s seminal notion of cooperativity can be modelled in such a setting
by assuming that the utilities of A’s decision problem are shared by all
conversational participants. However other notions at play in the Gricean
maxims have proven somewhat more resistant to formal explication. The key
target of decision-theoretic pragmatics is a formal representation of the Gricean
notion of relevance. [Mer99] gave an explication in terms of resolving a back-
ground question, while Prashant Parikh and Robert van Rooij have described
fully decision-theoretic models where comparative relevance corresponds to
effectiveness in resolving a decision problem [Par01; Roo03a; Roo03b].

4.1 · A measure of relevance

Van Rooij argues for a measure of relevance given by the value of sample
information (VSI), a notion from statistical decision theory (see e.g. [RS61]).
We will use this measure; I give now an argument that derives it.

Suppose the agent is facing a decision problem $D$, and receives the informa-
tion that $\varphi$. We will set awareness updates aside for the moment, so the new
decision problem can be written $D[\varphi]_b$ ($D$ transformed by a pure belief
update with $\varphi$, implemented by conditionalising on the support of $\varphi$).

\[\text{The observant reader will note that the references are to papers by “Robert van Rooy”. The spelling variation is a quirk of Dutch orthography. The modern Dutch alphabet contains a letter known as the ‘long y’, variously written $y$, $\text{"y}$ or $ij$ (the distinction between the latter two vanishes in handwritten cursive). Confusingly to the English-speaking reader, IJmuiden (a port close to Amsterdam) is so spelled (and capitalised!) but may be found alphabetised, in a Dutch index, just before Y. (Or it may not. Usage still varies a lot within the Netherlands, with even subject matter having some influence — telephone directories list Cruyff next to Cruyff, while modern dictionaries would split them, filing IJmuiden under I.) Such pedantry aside, Robert van Rooy and Robert van Rooij are the same person, who nowadays prefers the latter spelling.}\]
Intuitively, the value of the information that \( \phi \) should be something like the difference between the expected values of her decision problem before and after learning that \( \phi \). Suppose we implement this naïvely: the value of \( \phi \) is simply \( \text{val}(D[\phi]_b) - \text{val}(D) \) (recall that \( \text{val}(D) \) is the expected value of any best action in \( D \), in other words, the utility the agent can expect if she makes the best possible choice of action). But now ‘unwelcome information’ has negative value! If there is only a 0.01% chance that I have cancer, the expected utility of taking no treatment can be quite high (\( \text{val}(D) = 10 \), to name an entirely arbitrary figure); but if I learn that I do have cancer, my best action may be to take an expensive and painful chemotherapy treatment (\( \text{val}(D[\phi]_b) = 2 \), incorporating both the unpleasantness of the treatment and the chance that it nonetheless is unsuccessful). Because the news that I have cancer confirms that the state of the world is an inherently bad one, according to this measure the information has negative value; I would rather not learn that I have cancer, even though remaining in ignorance will lead me to take an action (not seeking treatment) leading to an unpleasant death (\( \text{BA}(D) = \{ \text{no treatment} \} \), but \( \text{EU}_{D[\phi]_b}(\text{no treatment}) = 0 \)).

The value of sample information instead recognises that if \( \phi \) is in fact the case, however unwelcome the fact may be, the information (that \( \phi \)) is never of negative value. On the standing assumption that the information (that \( \phi \)) is true, it establishes the fact that \( \phi \); the values to be compared are the expected utilities of acting with or without the information that \( \phi \), but on the standing assumption that (the fact that) \( \phi \) holds. The definition is as follows:

**Definition 5.11: Value of sample information (VSI).**

\[
\text{VSI}_D(\phi) = d \text{EU}_{D[\phi]_b}(\text{BA}(D)) - \text{EU}_{D[\phi]_b}(\text{BA}(D[\phi]_b)).
\]

That is, we compare two sets of actions: those that were considered best in the original decision problem (\( \text{BA}(D) \)), and those that are considered best after learning that \( \phi \) (\( \text{BA}(D[\phi]_b) \)). But we take the expected utilities of all of these actions with respect to the updated decision problem \( D[\phi]_b \) (note the subscripts on the expected utility calculations). VSI represents the value of receiving the information that \( \phi \), whereas the naïve measure above represented the value of bringing it about that \( \phi \); bringing it about that I have cancer has negative value, but if I do have cancer then learning that fact has positive value.

---

This unexpected consequence seems to have been overlooked in early work in the field (see for example [Par92] for an implicit use of the naïve notion in a game-theoretic setting). Probably this is because of an early focus on game-theoretic models of pure coordination, in which it is the joint behaviour of the players that largely determines the utility outcome, rather than the state of the world (there is usually no such thing as inherently bad news in such a model).
4.1.1 Some properties of interest

The first notable property of VSI has already been alluded to: it is non-negative. Information may be irrelevant (when $\text{VSI}_D(\varphi) = 0$) but never actively harmful. I will mention one apparent counterexample, in order to dismiss it (and a similar class of objections). Jim has a wife Jane and a mistress Joan, and neither of the two women know of the existence of the other. One day by coincidence both Jane and Joan happen to visit the same café. Now the information that Jane is in the café would lead Jim to go in himself, leading to a very unpleasant scene; surely this information should have negative value?

The argument relies on a refusal to take probabilistic expectations seriously enough. If Jim prefers to visit the café on the strength of the information that Jane is there (and taking into account the possibility that Joan might be as well), the chance that both women are there at the same time must be very small. It is that chance that the calculation makes reference to. The fact that the actual world happens to inhabit that low-probability region is simply a case of bad luck for Jim; over a large enough sample of alternative worlds this bad luck would ‘average out’ to approximate the expectation calculation.\footnote{See however Section 6.1 for a suggestion that this argument does not carry over to the case of awareness.}

It is easy to see that in such cases of ‘dangerous information’, more specific information that ‘resolves the danger’ will always have a higher value. In this case, both “Jane and Joan are in the café” and “Jane is in the café and Joan is not” have a higher value than “Jane is in the café”, in the respective worlds where they are true; the former because it leads to a more appropriate action, and the latter because it eliminates the small possibility that the action taken —entering the café— will lead to disaster. The situation is slightly different when $\varphi$ can be an uninformative utterance leading to an awareness update; I will return to the point in Section 6.1.

Another property of VSI is that it can be non-zero only if the set of best actions changes between $D$ and $D[\varphi]_b$. According to Van Rooij, “it doesn’t seem unnatural to say that a cooperative participant of the dialogue makes a relevant assertion in case he influences the action you are going to perform” [Roo03b, 735, orig. ital.]. On the other hand, the measure seems to be too strong, in that information can be intuitively relevant even if it does no more than confirm the optimality of an action taken under uncertainty. [Roo03b] acknowledges this point, but cites the naïve measure given above (also used by [Par01]) as a solution, which we have seen cannot be suitable. A more promising approach might acknowledge the uncertainty we have, as decision-makers, about the values of our probabilistic uncertainty. Given such higher-order uncertainty, small differences in expected utility (between two actions under consideration) are a risky basis for decision-making, while large differences are more robust and
reliable; I will not, however, introduce higher-order uncertainty into the model.

One final note on formalism, before we turn to the applications: it is because the best actions of $D$ are evaluated in $D[\varphi]_b$ that we need a definition for the average expected utility of a set of actions. By definition, the best actions of $D$ all have the same expected utility in $D$; however they may differ in effectiveness in the updated decision problem. The definition we have given corresponds intuitively to an agent who chooses randomly (with uniform distribution) between actions with the same expected utility. In reality such choices might also be influenced by other factors, in particular by considerations of risk (in the same way that in reality information that simply reinforces a decision may still be intuitively relevant). Again, I will not model this possibility; intuitively, however, the effect would seem to be to slightly raise the value of ‘extreme’ information favouring the risky alternatives.

4.2 · VSI for pragmatic reasoning

The use of VSI in reasoning is based on a presumption of relevance: the hearer assumes the speaker has followed the Gricean maxim of relevance, and tries to find an interpretation of the utterance that makes this the case. (It is thus unsuitable for deriving implicatures based on flouting the maxim, in which the semantic meaning is genuinely irrelevant.)

Consider the Gricean example “There is a garage around the corner”. Van Rooij writes [Roo03a, p. 1175] “Because B’s reaction can only resolve these issues [i.e., the decision problem “Where can I get petrol?”] when the garage is open, A understands that this is conversationally implied by B; otherwise the relevance [under VSI] of his assertion would be 0, i.e. his information would be pointless.” This argument is parallel to Grice’s own (although I will argue below that this example is better thought of as an awareness update) but the decision-theoretic framework allows some more subtle effects that are otherwise difficult to capture informally without ad hoc argumentation. Two examples (taken from two papers of Van Rooij) will suffice to give the flavour, although I will give only the most cursory presentation; for more examples and a more nuanced consideration of the possible complications see the respective publications.

4.2.1 · Generalised quantity implicatures [Roo03a]

A quantity implicature arises from Grice’s (first sub)maxim of quantity, “Make your contribution as informative as required for the current purposes of the exchange.” Early work on these implicatures focused on so-called implicational scales such as $\langle$some, most, all$\rangle$ and $\langle$or, and$\rangle$ [Hor72; Gaz79]. In a sentence containing a scalar term, such as “I drank some of the beer in the fridge”, replacing the term some with one higher on the same scale produces an informationally stronger statement: “I drank all of the beer in the fridge”, for instance. If informativity corresponds to entailment in the obvious way, then
using a term low on a scale implicates that the terms higher on the scale do not apply, since otherwise some alternative utterance would be more informative and the speaker should have used it.

[Roo03a] points out (pg. 1176) that this formulation, amenable though it is to formal treatment, ignores Grice’s original phrase “for the current purposes of the exchange”. The result is that some quantity implicatures cannot be treated in this way. Van Rooij cites the following example, due to [Hir85]:

(4) The setting is a job interview.
   a. Interviewer: Do you speak Portuguese?
   b. Applicant: My husband does.

The applicant’s answer clearly implicates that she does not speak Portuguese, by the quantity maxim in its original form. But “The applicant speaks Portuguese” and “The applicant’s husband speaks Portuguese” do not stand in any kind of entailment relation, so by strict standards of semantic informativity (underlying the systems of Horn and Gazdar) this conclusion cannot be drawn. VSI provides a generalisation of semantic informativity: it treats the scalar cases in the same way, but it also puts an ordering on utterances like these, that do not stand in any entailment relation with each other. The applicant’s utterance is less relevant, by VSI, than the alternative “I speak Portuguese”, and so by parallel reasoning to the scalar case it implicates that the alternative does not hold.

4.2.2 · Mention-some questions [Roo03b]

What counts as ‘fully answering a question’ depends not just on the semantics of the question and the answer, but also on the purposes with which it was asked. A classic distinction is between the questions “Who was at the party?” and “Where can I buy an Italian newspaper?” Naming a single newspaper stand is sufficient answer to the second question, while naming a single person who attended the party usually will not satisfy the person asking the first. It is not enough to draw a semantic distinction between partial and complete answers, because the same questions can be asked with the acceptability pattern of partial answers reversed:

(5) a. I’m making a survey of newspaper stand quality around the city. First question: where can I buy an Italian newspaper?
    b. I need to hear what happened from somebody who was there. Who was at the party?

What these additional sentences do is make explicit a particular decision problem the questioner is trying to solve. For some problems the VSI of a partial answer is just as high as that of a complete answer (if you tell me I can get an Italian newspaper at the station, I don’t care whether I can also get one
Chapter 5 · Pragmatics of decision-making

at the town hall), while for others the value rises with the specificity of the answer (“All right, John and Pete were at the party, but what about Joan and Petra, you haven’t told me about them yet.”).

The pragmatic relevance of these facts is the well-known phenomenon of **exhaustification** of answers. If I ask “Who (from our circle) was at the party?” and get the answer “John and Pete”, I will typically conclude from that “... and not Joan or Petra”, by the same kind of quantity-based reasoning mentioned above. In contrast, “Where can I buy an Italian newspaper” is a **mention-some** question which does not trigger such exhaustification: there is no temptation to augment the answer “At the station” with the quantity implicature “... and nowhere else”. If the inferences are the result of decision-theoretic relevance reasoning, then this phenomenon is entirely to be expected (as is the reversal in (5) above). What matters is not the semantic meaning **simpliciter**, but the influence of that meaning on the decision problem being solved; if more specific answers would be relevant but are not given, the speaker must not be in a position to give them (quantity implicature); alternative answers which would be **no more relevant**, however, as in the case of mention-some questions, have no influence and induce no implicatures.

4.3 · Calculating VSI with unawareness

The key to these analyses is to calculate the value of the utterance against the decision problem at hand, and then compare it to the value of putative alternatives. In just the same way, we can calculate the value of an utterance that leads not just to a change in the agent’s information but to a change in her awareness. The definition for the more general form is almost exactly the same: we simply remove the subscript for belief update ([ϕ] replaces [ϕ]₀), meaning that general updates with an awareness component are allowed. This small change has rather significant conceptual ramifications, though: we are no longer dealing with information (propositions whose truth in the actual world is assumed) but with epistemic change (which might not be appropriately evaluated in terms of truth and falsity). For this reason (which I will expand on in Section 6.1) [FJo08] called this measure the **value of epistemic change** (VEC): to emphasise the conceptual gulf between the two (formally almost identical) formulations.

**Definition 5.12**: Value of epistemic change (VEC).

\[
\text{VEC}_D(\varphi) = \text{d} \text{EU}_D[\varphi](\text{BA}(D)) - \text{EU}_D[\varphi](\text{BA}(D[\varphi])).
\]

The remarks about properties of VSI of course apply equally to VEC: an utterance is relevant (has non-zero VEC) just if it leads to a change in the set of best actions. For information this is nothing more than the standard account, but for pure awareness updates (triggered by **might**, questions, and so on) we
can note some interesting generalisations about the three kinds of updates given in (3).

One point to note in considering these examples is the joint epistemic status of the decision problem. The simplest possibility is that it is common knowledge between the speaker (who proposes the update) and the hearer (whose situation the decision problem directly represents). In that case the value of an utterance can be simple read off from the decision problem; however most likely the more common situation is when the speaker has imperfect knowledge of exactly which decision problem the hearer is trying to solve. In particular, speaker and hearer might have common knowledge of the probability distribution over states\textsuperscript{14} but the utilities of the hearer might be unknown to the speaker, at least in fine detail. This is particularly important because of the ‘hard-edged’ nature of the VEC calculation. As we will see, “You could take the bicycle” is only VEC-relevant if taking the bicycle is a best action; but then one might ask, why does the speaker not say “You should take the bicycle”? If she suffers some uncertainty about precisely which decision problem the hearer is trying to solve, however, the mild hedging becomes much more understandable. With this in mind, let us proceed to the examples.

4.3.1 · Did you think of taking the bicycle?
The most typical effect of this update on the decision problem will be nothing more than to broaden the range of actions available at each (proper) state, while the states themselves stay the same. It is easy to see that the only effect this could have on the set of best actions (if indeed the states are totally unchanged) is to add the action $b$ (take the bicycle) to that set (possibly even replacing the current best actions entirely). That is, as we started the chapter by noting, “Did you think of taking the bicycle?” is relevant according to this measure just in case taking the bicycle is a best action: the question can be interpreted as advice.

In line with this prediction, mentioning an action which obviously will not be taken is intuitively uncooperative behaviour: “You could rent a limousine,” or “Did you consider building an airship?” are not helpful contributions to a conversation about getting to work on time in the morning.

4.3.2 · There might be a tram strike
This update will either produce new states or split existing ones (overturning an assumption or increasing the finegrainedness of the agent’s beliefs). Overturning an assumption will only produce a change in the best action set if there is sufficient probability mass in the new state to exert an influence; “There might be a tram strike” is as irrelevant as “There might be a military coup

\textsuperscript{14}In itself a rather extreme abstraction, but the assumption of a common prior is widespread in the literature on multi-agent systems with probabilistic belief, since results are much harder to find without such constraining structure.
seizing control of the bus service” unless the prior probability of such a strike is relatively high. ‘Out of the blue’ the statement indeed seems irrelevant, but it would be quite reasonable in a climate of ongoing union disputes, or if the speaker knows that there will be a strike sometime this month but has forgotten exactly when. (Pragmatic reasoning based on a presumption of relevance leads to the implicature that a tram strike must be relatively likely, according to the speaker. We will come to this kind of reasoning in Section 5.)

According to the VEC measure, just splitting states can never be relevant. This is because the expected utility of an action in the state is nothing but a weighted average across the worlds in that state, which are the same worlds that would appear in the more finegrained states arrived at by splitting. As I mentioned above, though, a pure finegrainedness update only seems reasonable in a context where the distinction being drawn is indeed ‘small’ (in terms of outcomes and utilities); in this case a splitting update could be called ‘splitting hairs’ and seems intuitively irrelevant.  

4.3.3 · If you take the bicycle you might get a flat tyre

Assuming the consequent of the ‘outcome conditional’ \( a \rightarrow \phi \) is only possible if the action \( a \) is taken (as in this example), only the utility of \( a \) (among all the actions) can be affected by the update. That is to say, and remembering that non-zero VEC requires a change in the set of best actions, such an utterance must be an argument for or against the action being mentioned (‘for’ if it is not currently a best action, ‘against’ if it is). This certainly accords with intuition.

More problematic is the possibility that a relevant consequent is introduced by an irrelevant antecedent, as in “If you take the bicycle there might be a tram strike.” So long as the tram strike possibility is relevant (see above), this infelicitous conditional will get positive VEC. However I don’t think this is a problem for the account, per se: we just need more than a measure of relevance (even one that incorporates the quantity maxim) to correctly predict pragmatic felicity. In particular, the Gricean maxim of manner seems applicable here (since the relevance judgement rests entirely on the consequent, why is the utterance given in such an unnecessarily complicated manner?). It might also be possible to adapt standard Gricean accounts constraining the use of conditionals with unrelated antecedent and consequent to the awareness context.  

\(^{15}\)In a multi-agent setting with several conversational participants, a ‘hair-splitting’ update might be an invitation to another agent to share information: Ann distinguishes between \( p \) and not-\( p \) in the hopes that Bob will tell her which is the case. Such examples are only peripherally awareness-related, however, since questions exist for precisely this purpose. Awareness enters the picture only to establish the necessary condition for the question being asked, namely that the person asking it has considered the possibilities involved.

\(^{16}\)See for example [Frao, Chapter 5]. The potential difficulty is that such accounts tend to rely on (potential) informativity, which is no longer the only measure of relevance in the context of possible awareness updates.
4.4 · Unawareness and probabilities

One generalisation applies across all these updates, linking the probabilistic representation with the notion of awareness. I have argued for a notion of assumptions related to expectations of normality; this notion carries over even more strongly into the probabilistic setting. Our assumption-formation faculty (formally represented by the ordering generating the set $W_o$) and whatever faculty it is that draws our attention spontaneously to particular atomic formulae (not represented in the system, but implicit in every example in which the agent is already paying attention to certain concepts and not to others) works reasonably well most of the time. It must do, or we would not be able to form stable and reliable beliefs, and go about our daily business more or less successfully.

But if this is the case, then we would expect (as a broad generalisation, not a definite prediction for individual cases) the probability mass hidden by assumptions to be relatively small compared to that ‘visible’ within the window of awareness. If the actual world is quite likely to be one the agent does not entertain, then her assumption-formation faculty is not doing its job properly. In the Netherlands it is perfectly reasonable to assume that there is no tram strike or military coup in progress; it is evidence of a faulty assumption-formation faculty, on the other hand, to assume that it will not rain without checking a weather forecast.

This carries over to a generalisation about the relevance of awareness updates: when they apply to states of the world (rather than possible actions), the states they draw attention to should generally be associated with extreme utility values. This must be so if the probability mass they reveal is relatively small, since otherwise they stand little chance of changing the set of best actions. And indeed, this prediction too seems to be borne out. “If you cycle you might get a flat tyre” would typically be irrelevant (or, at most, an argument for taking a puncture repair kit), since the probability of a puncture on any particular trip is fairly low. If the agent is in a terrible hurry, though, and the time gained by leaving immediately on the bicycle rather than waiting five minutes for the next tram is significant, the extreme disutility of the puncture (and subsequent lost time) might overturn such a judgement.

5 · Relevance reasoning

So far we have only considered evaluating the relevance of utterances ‘from the outside’, as observers. Our agents too, though, can perform such calculations, and these can form the basis for nontrivial pragmatic reasoning. The general schema is that the hearer assumes the speaker to produce a relevant utterance (by the standard of VEC). If the utterance would only be relevant on the assumption that the speaker possesses some special knowledge, then the hearer concludes that she does indeed possess that knowledge (and typically that she
intended to communicate it). Here is an example.

**Example 5.13:** Bob the Baker. Bob (*who is an expert baker*) is visiting his friend Farmer Pickles (*who isn’t*).

**Pickles:** I was going to bake a cake but I haven’t got any eggs!

**Bob:** Did you think of making shortbread instead?

**Pickles:** I didn’t, in fact I didn’t even know that you don’t need eggs to make shortbread! Thanks, Bob!

Remember that raising the possibility of an action can only be an argument *for* that action, under the relevance criterion of VEC. If Bob is advising Pickles to make shortbread, this must be because Bob believes that the recipe for shortbread does not require eggs (since otherwise trying to bake shortbread would be no better than trying to bake a cake, and certainly worse than not baking anything at all). But since Bob is an expert baker (a competence assumption, in the terminology of [RS04]), his belief about shortbread can be taken to be factual.

We have come a long way from the relevance measure of VSI. Bob gives no overt information at all: his utterance is a question, which might indeed alter the common ground (by ‘raising issues’, or delineating his interest in different potential answers) but certainly does not do so by eliminating worlds. The immediate effect of the question is rather to *add* possibilities, at least in the sense of possible actions — a *reduction* in information as standardly conceived. And yet, by pragmatic reasoning Pickles indeed gains just the information he needs.

Indeed, I would argue that Grice’s famous petrol example belongs more in this context than as an example of reasoning from a purely informative utterance. “There is a garage around the corner” is very likely a proposition the car owner was not attending to (otherwise, parallel to the situation with Walt and the car keys, she should be looking around the corner to check). The action “Go around the corner and get petrol from the garage” is equally unlikely to be a part of the original decision problem, but once it has been incorporated the reasoning is entirely parallel.

The example highlights one deficiency of the propositional approach: awareness of the action “Go around the corner and get petrol from the garage” is clearly closely related to awareness of the garage being around the corner, a relation which is entirely obscured by representing them as distinct proposition letters. A first-order model in which the agent may be unaware of actions, properties, and objects and in which these combine in the natural ways would be much more appropriate for such examples; I will consider the beginnings of such a theory in the next chapter. Here is a third example on the same structural lines, due to Anton Benz, which requires a similar approach.
Example 5.14: Travel expenses. I am to submit a form requesting repayment of travel expenses to the administrators of the zas. I ask at the front desk where I should go, and am told, ‘Mrs Schmidt is in room 2.15.’

The reasoning that leads me to deliver my expense form to Mrs Schmidt is exactly parallel to that leading Grice’s stranded driver to the garage for petrol. However Benz’s example has a further symmetry property which causes difficulties for standard probabilistic accounts, but which can be solved using unawareness.

5.1 · A digression on symmetry

An awareness-based account does better than a purely propositional account for this example because we can have actions ‘offstage’ for the agent, so that the information that Mrs Schmidt is in a particular room brings with it the implicature that a particular action is optimal. However Benz was concerned with a refinement of the example, which raises even tougher questions about the use of possible-world semantics for such problems:

Example 5.15: Symmetrical travel expenses (Benz). As before, I am to submit a travel expense form at the zas. This time I know that it must go to either Mrs Schmidt or Mr Müller, and that their rooms are 2.15 and 2.16; I don’t know, though, who has which room, nor do I know who should get the form. As before, the front desk assistant tells me, “Mrs Schmidt is in room 2.15.”

The problem for standard theories is the extreme symmetry of the example. Given my background knowledge, the proposition expressed by “Mrs Schmidt is in room 2.15” is identical to the proposition expressed by “Mr Müller is in room 2.16”. Under such (admittedly artificial) conditions we have no formal representation of the intuitive ‘aboutness’ of the two utterances (that the first is about Mrs Schmidt and room 2.15, while the second is about Mr Müller and room 2.16), since both pick out exactly the same subset of my belief set. But of course, this is precisely what a theory of awareness gives us: an utterance is not reduced to its propositional information, but is (in effect) a pair consisting of its propositional meaning and the things it ‘talks about’ (the atomic formula as a syntactic object, or in our informal extension for this example, the people and places it mentions).

That is not enough to magically solve all our difficulties, however. It is hard to imagine how I could know that Mrs Schimdt and Mr Müller can be found in rooms 2.15 and 2.16 (although in which order I am uncertain) without already being aware of the two people and places. The awareness update produced by “Mrs Schmidt is in room 2.15”, then, is vacuous, and it seems we are left in the same quandary as the theories without awareness.

There is, however, a loophole: pragmatic reasoning. The attention update
has no effect, but it may have been intended to have an effect. Specifically, it may have been intended to make me aware of Mrs Schmidt just as in the first example, to send me to room 2.15 to hand in my travel expense form. This would be the case if the speaker thought I was unaware of Schmidt and Müller, which seems reasonable, but in fact it also holds for all deeper nestings of uncertainty. In all epistemic situations except common knowledge that I am aware of Mrs Schmidt (that is, any finite-depth mutual certainty followed by uncertainty), a pragmatic case can be made for using “Mrs Schmidt is in room 2.15” to indicate that that is where I should go. In contrast, no pragmatic case can be made for using the same statement to direct attention to Mr Müller in any configuration of mutual belief or uncertainty. In any case but common knowledge, there exists the possibility that the assistant uses the utterance to send me to Mrs Schmidt; since this is the only possibility that makes the utterance relevant (it can never draw attention to Mr Müller), the mere existence of such a possibility makes it the preferred interpretation.

I don’t mean to claim that we go through such complex reasoning in considering such simple examples. Rather, I suggest that this phenomenon lies at the functional root of a far simpler notion: the influence of aboutness on Grice’s maxim of manner. The possibility of awareness confusions, even if only in rather extreme cases, justifies a manner ‘submaxim’: in order to convey information about \(x\), use (preferentially) an expression about \(x\). This is of course nothing but common sense, and it takes an artificial example like the one we are considering here to force us to acknowledge that our formal systems, elegant though they may be, do not represent this level of common sense behaviour. To this extent the awareness model is an improvement: it tells us why our common sense behaves the way it does, if not necessarily (in this case) how.

5.1.1 · Horn’s division

Unawareness can help solve such symmetry problems because it is inherently asymmetric. An agent aware of \(p\) can understand the viewpoint of unawareness of \(p\), but the converse is impossible. So in the previous example, despite the semantic equivalence (given the background information) of the information about Schmidt and Müller, whichever one is not mentioned is ‘invisible’ to the agent.

Another application of this idea is to Horn’s division of pragmatic
Labour: the rule that when marked and unmarked expressions have the same semantic meaning, the marked expression is used for non-stereotypical cases of the meaning while the unmarked expression goes with stereotypical cases (the notion, although not the term, comes from [Hor84]). This is another case where game theory has produced effective but perhaps over-complicated explanations; [Roo04] shows that Horn’s division emerges from certain kinds of evolutionary models (although an equally important part of the paper is given over to game-theoretic models that do not produce Horn’s division, in some cases quite counter to first expectations).

Van Rooij also discusses static (ahistorical) game-theoretic models, but concludes that they do not suffice to explain the Horn pattern, for reasons that have a lot to do (again) with symmetry. Without getting into the details, the competing possible strategies produced by systematically permuting utterances and meanings have many structural properties in common with the Horn strategy (even though the Horn strategy is in the end the most efficient at the global level), so that standard game-theoretic techniques cannot ‘see the differences’ between them. Certain kinds of evolutionary model, on the other hand, allow efficiency to influence the long-term behaviour of the model; since the historical dimension is a crucial part of this process, Van Rooij concludes that Horn’s division is (in its general form) a convention of the language, stabilised over the long term by its relative efficiency.

The same result can, however, be derived using unawareness under simpler (and arguably more natural) assumptions. Gricean pragmatic reasoning proceeds by comparisons between the utterance actually made and the possible alternatives the speaker could have said. If the awareness relations between these are systematically asymmetric, no appeal to conventions or long-term linguistic evolution need be made.

Suppose that the marked form is one the agent is typically unaware of, and that the marked form triggers awareness of the unmarked form but not vice versa. Van Rooij begins his paper with a standard example: “Miss X produced a series of sounds that corresponded closely with the score of ‘Home Sweet Home’.” After puzzling over this for a moment, you probably realised that she sang the song (this insight sets in train the pragmatic reasoning that starts with “Why didn’t he just say so then?”). If he had said instead that Miss X sang ‘Home Sweet Home’, nothing would have drawn your attention to the possible alternative utterance “Miss X produced a series of sounds…”.

The intuitive appeal of this account is its asymmetry. The usual case (unmarked form, stereotypical meaning) involves unawareness of both the marked utterance in its syntactic form and the very dimension of variation on which the stereotypical and unexpected events differ (that is, whether the singing was good or bad). From the perspective of awareness of the marked
form, we can still ‘see’ this alternative possibility; this corresponds to the intuitive question “If she just sang (in an ordinary way), why didn’t he just say so?” Searching for a motivation, we must spontaneously come to attend to the quality of the singing as a possible additional variable.\textsuperscript{18} Only after attending to the possibility of bad singing do we realise explicitly that our stereotypical assumption about singing is that it is pleasant.

This means that the account differs substantially from Van Rooij’s treatment of ‘stereotypicality’. For him the stereotypical event is the one with highest probability (this is mathematically required for Horn’s division to be the most efficient long-term strategy). While generally appealing, this idea runs into difficulties with specific cases. What if, for instance, Miss X is always a terrible singer? Then we would be forced (under the explicit terms of Van Rooij’s account) to conclude from the marked expression that she sang well, unless we fall back on an ad hoc procedure for generalising probabilities that relates the stereotypical interpretation of the (specific) utterance about Miss X to knowledge about singers in general. That stereotypical singing is pleasant singing is, I would suggest, a fact about awareness rather than a fact about probabilities. (The lowest singing-worlds in our ordering tend to be pleasant-singing ones.) This fact is connected to frequencies of events, in that our awareness orderings must behave roughly consistently with probabilistic expectations, but it does not reduce to them. And, in particular, if we are not attending to the question of quality, we may very well hold a general assumption (that the singing was pleasant) which we would repudiate when the question is explicitly raised (bearing in mind what we also know about Miss X).

6 · Summary

The treatment of Horn’s division of pragmatic labour shows one way in which unawareness can aid formal pragmatics: in simplifying the representations of core phenomena, reducing complex game-theoretic explanations to more tractable decision theory. More generally, awareness can reduce the unwanted effects of symmetry, whether probabilistic and numerical (as in the case of Benz’s expense form) or structural (in the game-theoretic representation of Horn’s division).

A second achievement is the extension of standard techniques (such as relevance reasoning from VSI) beyond the standard cases (the shift from informative to allusive dialogue). Formal pragmatics has had its greatest successes in reasoning based on simple informative language use, since semantic meaning gives an immediate formal handle with which to manipulate utterances in rea-

\textsuperscript{18}My account is silent on how this particular variable is arrived at; but so indeed are all others that I am aware of. Game-theoretic approaches assume that the possible meaning space is already given, so that the only possibilities are “sang badly” and “sang well” (why not also “whistled”, for instance?). Informal accounts of course fare no better, on what is essentially an awareness question.
soning. Unawareness provides one way in which the ‘tyranny of informativity’ can be broken; certainly this is only a small part of non-informative language use, but any movement in this direction can only be to the benefit of the field. In particular, unawareness models explain how questions, statements with might, and similar paradigms of uninformativity can be pragmatically interpreted as advice, against all semantic expectation.

In the other direction, decision-theoretic models have told us something interesting about the pragmatic conditions under which drawing attention to new possibilities counts as cooperative language use. In particular, the requirement that an awareness update make a difference to the actions the agent takes as best leads to a general condition on the relation between assumptions and probabilities: if an assumption is overturned then either it was concealing significant probabilistic weight (the ‘head-slap moment’ that greets Pat’s helpful suggestion to Walt) or, in the more usual case where the concealed probability mass is relatively small, the utilities of outcomes involved must be extreme (whether extremely large or extremely small).

Let me finish by mentioning a persistent loose end, which raises questions about the interpretation of the notion of ‘information’ in models with awareness.

6.1 · Some speculation: hurting attention?

The problem is that not all of the arguments for VSI transfer immediately to the unawareness context. In particular, while I am convinced that information must have a non-negative value (and that putative counterexamples rest on an insufficiently serious interpretation of probabilistic expectations), the parallel argument for attention seems to me rather weaker.

The main difference is that information converges in the limit, while unawareness updates can broaden the agent’s horizons apparently without restraint. Information has non-negative value because each informative update brings the agent closer to the fixpoint of no uncertainty (a belief state containing only the actual world); but each awareness update takes her further away from such a fixpoint, by introducing new possibilities.

Consider again the unfortunate case of Jim and Joan and Jane. A malevolent speaker who saw Jane and Joan in the café could manoeuvre Jim into making a wrong decision, by saying “Joan is in the café”. But there are limits to his potential for mischief: the more information he chooses to give, the more appropriate Jim’s reaction will be, so that in the limit Jim knows which the actual world is and takes the action maximising (not expected but actual) utility. A malevolent speaker in the system with awareness, on the other hand, can cause unending misery, because outré possibilities can always be found that are highly unlikely but that carry extreme utilities.

This brings us to considerations of scepticism, which I am not sure form a genuine continuum with mundane awareness data (see the discussion in
Chapter 7 Section 2.1). In any case the pragmatic reasoning we have been concerned with in this chapter makes the Gricean assumption of cooperation, so the problematic consequence of malignancy is irrelevant; I wonder, though, whether similar problems will arise in a setting of only partial coordination (see the treatment of credibility in [Fra09]) and what the consequences for an awareness-relative measure of relevance should be.