How to Set Up Normal Optimal Answer Models

Anton Benz

Zentrum für Allgemeine Sprachwissenschaft, Berlin

Abstract. We investigate the role of multi-attribute utility analyses in game theoretic models of Gricean pragmatics, i.e. for finding a model of the linguistic context of an utterance and for the calculation of implicatures. We investigate especially relevance implicatures of direct answers. The work is based on the optimal answer model (Benz, 2006; Benz & v. Rooij, 2007). We argue that multi-attribute utility functions play an essential role in finding the appropriate models. We concentrate especially on default assumptions which are necessary in order to calculate the correct implicatures. As normality assumptions play a central role in the construction of these models, we call them *normal* optimal answer models. We introduce rules which provide guidelines for setting up these models.

1 Introduction

This paper can be characterised as a preparatory study for a theory of relevance implicatures in discourse. In order to set up such a theory, we need to know how to construct a game theoretic model of an utterance situation from a given discourse. We argue that multi-attribute utility functions play an essential role in this process. The core examples are of the following form:

(1) Peter: I have to buy wine for our dinner banquet. I will get into trouble with our secretary if I spend too much money on it. We still have some white wine. Where can I buy red wine?
Bob: At the Wine Centre.
+> Peter can buy red wine at a low price at the Wine Centre.

The root question in (1) makes no reference to the price of wine. Nevertheless, the contextually given objective of buying wine at a low price has an impact on the implicature of the answer. These objectives are provided by the linguistic context which stands in a background relation to the question. In (1), we find a domain object d, the wine shop, and two relevant attributes, $S(\, . \,)$ for *selling wine*, and $L(.)$ for *selling at low prices*. In addition, the inquirer can just perform a random search with expected utility $0 < \varepsilon < 1$. The essential parameters of the answering situation are shown in Table 1. Going to shop d is an optimal choice in world w_1 only. Hence, the inquirer can infer from the recommendation to go to d that the actual world is w_1.

An obvious question is, how do we arrive at the game theoretic model starting from the discourse in (1)? This question involves first of all the question how to

Ω	$S(d)$	$L(d)$	search	$Good(d)$
w_1	1	1	ε	yes
w_2	1	0	ε	no
w_3	0	1	ε	no
w_4	0	0	ε	no

Table 1. The table for Example (1).

determine the set of possible worlds Ω. Then, we have to ask, how do we know the speakers and hearers knowledge about Ω? How do we know their utilities over outcomes and the actions between which the hearer can choose? Only part of the answers to these questions is explicitly stated in (1). In this paper, we propose a procedure for constructing a game theoretic model from a discourse which provides answers to these questions. This procedure is based on a multi-attribute utility analysis and some default rules. For example, the set Ω will be constructed as the set of all attribute-value assignments, and the default rules introduce assumptions about the independence of elementary events and the even distribution of probabilities.

The paper divides into three parts. First, we will introduce the game theoretic framework on which we base our analysis. This is the Optimal Answer (OA) model. The version presented here is slightly more general than (Benz, 2006; Benz & v. Rooij, 2007). The second part, Section 3, contains the main body of results. We introduce the multi-attribute utility theory and use it for analysing an extended example. Based on this analysis, we generalise the procedure used there and introduce a general prescript for constructing optimal answer models. Finally, in the last section, we show how to use normal optimal answer models for deriving systematic classifications of dialogue situations and calculating the associated lists of optimal answers. In particular, we provide a complete list of all situations in which the addressee has to choose a domain object based on preferences defined by two attributes.

2 The optimal-answer model

Grice (1989, p. 26) characterised conversation as a *cooperative effort*. The contributions of the interlocutors are not isolated sentences but normally subordinated to a joint purpose. In this paper, we will always assume that questioning and answering is embedded in a decision problem in which the inquirer has to make a choice between a given set of actions. His choice of action depends on his preferences regarding their outcomes and his knowledge about the world. The answer helps the inquirer in making his choice. The quality of an answer depends on the action to which it will lead. The answer is optimal if it induces the inquirer to choose an optimal action. We model answering situations as two-player games. We call the player who answers the *expert* E, and the player who receives the answer the *inquirer* I.

For Grice, the information communicated by an answer divides into two parts, the semantic meaning of the answer and its implicated meaning. In our definition of *implicature*, which we provide later, we closely follow Grice's original idea that implicatures arise from the additional information that an utterance provides about the state of the speaker:

> "... what is implicated is what is required that one assume a speaker to think in order to preserve the assumption that he is observing the Cooperative Principle (and perhaps some conversational maxims as well), ..." (Grice, 1989, p. 86)

In a game theoretic model, what the speaker utters is determined by his strategy s, i.e. a function that selects a sentence for each of his possible information states. When the inquirer receives answer F, then he knows that the expert must have been in a state K which is an element of $s^{-1}(F) = \{K \mid s(K) = F\}$, i.e. the set of all states which are mapped to F by s. Lewis (2002, p. 144) calls this the *indicated* meaning of a signal F. We identify the implicature of an utterance with this indicated information. This identification implies that, once we know s, the implicatures can be calculated. Hence, all depends on how we can know the speaker's strategy s. This knowledge will be provided by the Optimal-Answer (OA) Model and its later modifications.

2.1 Optimal answers

The OA model tells us which answer a rational language user will choose given the inquirer's decision problem and his own knowledge about the world. Instead of introducing full signalling games (Lewis, 2002), we reduce our models to the cognitively relevant parameters of an answering situation. We call these simplified models *support problems*. They consist of the inquirer's decision problem and the answering expert's expectations about the world. They incorporate the *Cooperative Principle*, the maxim of *Quality*, and a method for finding optimal strategies which replaces the maxims of *Quantity* and *Relevance*. In this section, we ignore the maxim of *Manner*.

A decision problem consists of a set Ω of the possible states of the world, the decision maker's expectations about the world, a set of actions $\mathcal{A}$ he can choose from, and his preferences regarding their outcomes. We always assume that Ω is finite. We represent an agent's expectations about the world by a probability distribution over Ω, i.e. a real valued function $P : \Omega \to \mathbf{R}$ with the following properties: (1) $P(v) \geq 0$ for all $v \in \Omega$ and (2) $\sum_{v \in \Omega} P(v) = 1$. For sets $A \subseteq \Omega$ it is $P(A) = \sum_{v \in A} P(v)$. The pair (Ω, P) is called a finite *probability space*. An agent's preferences regarding outcomes of actions are represented by a real valued function over action-world pairs. We collect these elements in the following structure:

Definition 1 *A* decision problem *is a triple* $\langle (\Omega, P), \mathcal{A}, u \rangle$ *such that* (Ω, P) *is a finite probability space,* $\mathcal{A}$ *a finite, non-empty set and* $u : \Omega \times \mathcal{A} \to \mathbf{R}$ *a function.* $\mathcal{A}$ *is called the* action *set, and its elements* actions*;* u *is called a* payoff *or* utility *function.*

In the following, a decision problem $\langle(\Omega, P), \mathcal{A}, u\rangle$ represents the inquirer's situation before receiving information from an answering expert. We will assume that this problem is common knowledge. How do we find a solution to a decision problem? It is standard to assume that rational agents try to maximise their expected utilities. The *expected utility* of an action a is defined by:

$$EU(a) = \sum_{v \in \Omega} P(v) \times u(v, a). \tag{1}$$

The expected utility of actions may change if the decision maker learns new information. To determine this change of expected utility, we first have to know how learning new information affects the inquirer's beliefs. In probability theory the result of learning a proposition A is modelled by *conditional probabilities*. Let H be any proposition and A the newly learned proposition. Then, the probability of H *given* A, written $P(H|A)$, is defined as

$$P(H|A) := P(H \cap A)/P(A) \text{ for } P(A) \neq 0. \tag{2}$$

In terms of this conditional probability function, the *expected utility after learning* A is defined as

$$EU(a|A) = \sum_{v \in \Omega} P(v|A) \times u(v, a). \tag{3}$$

I will choose the action which maximises his expected utilities after learning A, i.e. he will only choose actions a where $EU(a|A)$ is maximal. We assume that I's decision does not depend on what he believes that the answering expert believes. We denote the set of actions with maximal expected utility by $\mathcal{B}(A)$, i.e.

$$\mathcal{B}(A) := \{a \in \mathcal{A} \mid \forall b \in \mathcal{A}\ EU_I(b|A) \leq EU_I(a|A)\}. \tag{4}$$

The decision problem represents the inquirer's situation. In order to get a model of the questioning and answering situation, we have to add a representation of the answering expert's information state. We identify it with a probability distribution P_E that represents his expectations about the world:

Definition 2 *A five-tuple* $\sigma = \langle \Omega, P_E, P_I, \mathcal{A}, u\rangle$ *is a* support problem *if* (Ω, P_E) *is a finite probability space and* $D_\sigma = \langle(\Omega, P_I), \mathcal{A}, u\rangle$ *a decision problem such that there exists a probability distribution* P *on* Ω*, and sets* $K_E \subseteq K_I \subseteq \Omega$ *for which* $P_E(X) = P(X|K_E)$ *and* $P_I(X) = P(X|K_I)$.

The last condition says that P_E and P_I are derived from a common prior P by Bayesian update. It entails:

$$\forall X \subseteq \Omega\ P_E(X) = P_I(X|K_E). \tag{5}$$

This condition allows us to identify the *common ground* in conversation with the addressee's expectations about the domain Ω, i.e. with P_I. The speaker knows the addressee's information state and is at least as well informed about Ω. Hence, the assumption is a probabilistic equivalent to the assumption about common ground that implicitly underlies dynamic semantics (Groenendijk & Stockhof, 1991). Furthermore, condition (5) implies that the expert's beliefs cannot contradict the inquirer's expectations, i.e. for $A, B \subseteq \Omega$: $P_E(A) = 1 \Rightarrow P_I(A) > 0$.

The expert E's task is to provide information that is optimally suited to support I in his decision problem. Hence, we find two successive decision problems, in which the first problem is E's problem to choose an answers. The utility of the answer depends on how it influences I's final choice:

Inquirer I Asks		Expert E answers		I decides for action		Evaluation
$\downarrow$		$\downarrow$		$\downarrow$		$\downarrow$
$\bullet$	$\xrightarrow{Q?}$	$\bullet$	$\xrightarrow{A}$	$\bullet$	$\xrightarrow{a}$	$\bullet$
$\uparrow$		$\uparrow$		$\uparrow$		$\uparrow$
expectations of I $\langle \Omega, P_I \rangle$		expectations of E $\langle \Omega, P_E \rangle$		expectations of I $\langle \Omega, P_I(\,.\, \vert A) \rangle$		utility measure $u(v, a)$

We assume that E is fully cooperative and wants to maximise I's final success; i.e. E's payoff, is identical with I's. This is our representation of Grice's *Cooperative Principle*. E has to choose an answer that induces I to choose an action that maximises their common payoff. In general, there may exist several equally optimal actions $a \in \mathcal{B}(A)$ which I may choose. Hence, the expected utility of an answer depends on the probability with which I will choose the different actions. We can assume that this probability is given by a probability measure $h(.|A)$ on $\mathcal{A}$. Then, the expected utility of an answer A is defined by:

$$EU_E(A) := \sum_{a \in \mathcal{B}(A)} h(a|A) \times EU_E(a). \tag{6}$$

We add here a further Gricean maxim, the *Maxim of Quality*. We call an answer A *admissible* if $P_E(A) = 1$. The Maxim of Quality is represented by the assumption that the expert E does only give admissible answers. This means that he believes them to be *true*. For a support problem $\sigma = \langle \Omega, P_E, P_I, \mathcal{A}, u \rangle$ we set:

$$\mathrm{Adm}_\sigma := \{A \subseteq \Omega \mid P_E(A) = 1\} \tag{7}$$

Hence, the set of optimal answers in σ is given by:

$$\mathrm{Op}_\sigma := \{A \in \mathrm{Adm}_\sigma \mid \forall B \in \mathrm{Adm}_\sigma \; EU_E(B) \leq EU_E(A)\}. \tag{8}$$

We write Op_σ^h if we want to make the dependency of Op on h explicit. Op_σ is the set of *optimal answers* for the support problem σ. As answers are propositions in our model, i.e. sets $A \subseteq \Omega$, it trivially follows that all propositions can be expressed.

The *behaviour* of interlocutors can be modelled by *strategies*. A strategy is a function which tells us for each information state of an agent which actions he may choose. It is not necessary that a strategy picks out a unique action for each information state. A *mixed* strategy is a strategy which chooses actions with certain probabilities. The hearer strategy $h(.|A)$ is an example of a mixed strategy. We define a (mixed) strategy pair for a support problem σ to be a pair (s,h) such that s is a probability distribution over $\mathcal{P}(\Omega)$ and $h(.|A)$ a probability distribution over $\mathcal{A}$.

We may call a strategy pair (s,h) a *solution* to σ iff $h(.|A)$ is a probability distribution over $\mathcal{B}(A)$, and s a probability distribution over Op^h_σ. In general, the solution to a support problem is not uniquely defined. Therefore, we introduce the notion of the *canonical* solution.

Definition 3 *Let* $\sigma = \langle \Omega, P_E, P_I, \mathcal{A}, u\rangle$ *be a support problem. The* canonical solution *to* σ *is a pair* (S,H) *of mixed strategies which satisfy:*

$$S(A) = \begin{cases} |\mathrm{Op}_\sigma|^{-1},\ A \in \mathrm{Op}_\sigma \\ 0 \ \textit{otherwise} \end{cases},\ H(a|A) = \begin{cases} |\mathcal{B}(A)|^{-1},\ a \in \mathcal{B}(A) \\ 0 \ \textit{otherwise} \end{cases}. \tag{9}$$

We write $S(\,.\,|\sigma)$ if S is a function that maps each $\sigma \in \mathcal{S}$ to the speaker's part of the canonical solution, and $H(\,.\,|D_\sigma)$ if H is a function that maps the associated decision problem D_σ to the hearer's part of the canonical solution. From now on, we will always assume that speaker and hearer follow the canonical strategies $S(\,.\,|\sigma)$ and $H(\,.\,|D_\sigma)$. We make this assumption because it is convenient to have a unique solution to a support problem; the only property that we really need in the following proofs is that $H(a|A) > 0 \Leftrightarrow a \in \mathcal{B}(A)$ and $S(A|\sigma) > 0 \Leftrightarrow A \in \mathrm{Op}_\sigma$.

The expert may always answer everything he knows, i.e. he may answer $K_E := \{v \in \Omega \,|\, P_E(v) > 0\}$. From condition (5) it trivially follows that $\mathcal{B}(K_E) = \{a \in \mathcal{A} \,|\, \forall b \in \mathcal{A}\ EU_E(b) \leq EU_E(a)\}$. If expert and inquirer follow the canonical solution, then:

$$\mathrm{Op}_\sigma = \{A \in \mathrm{Adm} \mid \mathcal{B}(A) \subseteq \mathcal{B}(K_E)\}. \tag{10}$$

In order to show (10), let $A \in \mathrm{Adm}$ and $\alpha := \max\{EU_E(a) \mid a \in \mathcal{A}\}$. For $a \in \mathcal{B}(A) \setminus \mathcal{B}(K_E)$ it holds by definition that $EU_E(a) < \alpha$ and $H(a|A) > 0$. $EU_E(A)$ is the sum of all $H(a|A) \times EU_E(a)$. If $\mathcal{B}(A) \not\subseteq \mathcal{B}(K_E)$, then this sum divides into the sum over all $a \in \mathcal{B}(A)\setminus\mathcal{B}(K_E)$ and all $a \in \mathcal{B}(A)\cap\mathcal{B}(K_E)$. Hence, $EU_E(A) < \alpha$, and therefore $A \notin \mathrm{Op}_\sigma$.

If $\mathcal{B}(A) \not\subseteq \mathcal{B}(K_E)$, then the answering expert knows that answering A would induce the addressee to choose a sub-optimal action with positive probability. Hence, we can call an answer A *misleading* if $\mathcal{B}(A) \not\subseteq \mathcal{B}(K_E)$; then, (10) implies that Op_σ is the set of all non-misleading answers.

2.2 Implicatures of optimal answers

An implicature of an utterance is a proposition which is implied by the assumption that the speaker is cooperative and observes the conversational maxims. More precisely, Grice linked implicatures to what the hearer learns from the utterance about the speaker's knowledge. The speaker's canonical solution maps his possible information states to utterances. Hence, the hearer can use this strategy to calculate what the speaker must have known when making his utterance. As the canonical solution is a solution, it also incorporates the information that the speaker is cooperative and follows the maxims.

We treat all implicatures as particularised implicatures, i.e. as implicatures that follow immediately from the maxims and the particular circumstances of the utterance context. The answering expert knows a proposition H in a situation σ iff $P_E^\sigma(H) = 1$. Hence, if the inquirer wants to know what the speaker knew when answering that A, he can check all his epistemically possible support problems for what the speaker believes in them. If σ is the support problem which represents the actual answering situation, then all support problems $\hat{\sigma}$ with the same decision problem D_σ are indiscernible for the inquirer. Hence, the inquirer knows that the speaker believed that H when making his utterance A, iff the speaker believes that H in all indiscernible support problems in which A is an optimal answer. This leads to the following definition:

Definition 4 (Implicature) *Let $\mathcal{S}$ be a given set of support problems with joint decision problem $\langle(\Omega, P_I), \mathcal{A}, u\rangle$. Let $\sigma \in \mathcal{S}$, $A, H \subseteq \Omega$ be two propositions with $A \in \mathrm{Op}_\sigma$. Then we set:*

$$A +\!\!>_\sigma H \Leftrightarrow \forall \hat{\sigma} \in \mathcal{S}\, (A \in \mathrm{Op}_{\hat{\sigma}} \rightarrow P_E^{\hat{\sigma}}(H) = 1), \tag{11}$$

If $A +\!\!>_\sigma H$, we say that the utterance of A implicates that H in σ*.*

The definition entails:

$$\text{If } A \subseteq H, \text{ then } A +\!\!>_\sigma H. \tag{12}$$

Hence, our use of '*implicate*' deviates from the common usage in which implicated information is something extra in addition to entailed information. In our usage, entailed information is part of the implicated information. This is just a matter of convenience. If no confusion can arise, we simply drop the subscript σ in $+\!\!>_\sigma$.

As the hearer has to check all support problems in $\mathcal{S}$, it follows that we arrive at the more implicatures the smaller $\mathcal{S}$ becomes. If $\mathcal{S} = \{\sigma\}$ and $A \in \mathrm{Op}_\sigma$, then A will implicate everything the speaker knows. The other extreme is the case in which answers implicate only what they entail. We show in Proposition 7 that this case can occur.

We are interested in cases in which the speaker is a real expert. If he is an expert, then we can show that there is a very simple criterion for calculating implicatures. We can call the speaker an expert if he knows the actual world; but we will see that a weaker condition is sufficient for our purposes. To make precise what we mean by expert, we introduce another important notion, the set $O(a)$ of all worlds in which an action a is optimal:

$$O(a) := \{w \in \Omega \mid \forall b \in \mathcal{A}\, u(w, a) \geq u(w, b)\}. \tag{13}$$

We say that the answering person is an expert for a decision problem if there is an action which is an optimal action in all his epistemically possible worlds. We represent this information in $\mathcal{S}$:

Definition 5 (Expert) *Let $\mathcal{S}$ be a set of support problems with joint decision problem $\langle(\Omega, P_I), \mathcal{A}, u\rangle$. Then we call E an* expert *in a support problem σ if $\exists a \in \mathcal{A}\, P_E^\sigma(O(a)) = 1$. He is an expert in $\mathcal{S}$, if he is an expert in every $\sigma \in \mathcal{S}$.*

This leads us to the following criterion for implicatures:

Lemma 6 *Let $\mathcal{S}$ be a set of support problems with joint decision problem $\langle(\Omega, P_I), \mathcal{A}, u\rangle$. Assume furthermore that E is an expert for every $\sigma \in \mathcal{S}$ and that $\forall v \in \Omega\, \exists \sigma \in \mathcal{S}\, P_E^\sigma(v) = 1$. Let $\sigma \in \mathcal{S}$ and $A, H \subseteq \Omega$ be two propositions with $A \in \mathrm{Op}_\sigma$. Then, with $A^* := \{v \in \Omega \mid P_I(v) > 0\}$, it holds that:*

$$A \mathrel{+\!>} H \text{ iff } A^* \cap \bigcap_{a \in \mathcal{B}(A)} O(a) \subseteq H. \tag{14}$$

Proof. We first show that

$$(\exists a \in \mathcal{A}\, P_E^\sigma(O(a)) = 1 \;\&\; A \in \mathrm{Op}_\sigma) \Rightarrow \forall a \in \mathcal{B}(A):\; P_E^\sigma(O(a)) = 1. \tag{15}$$

Let a, b be such that $P_E^\sigma(O(a)) = 1$ and $P_E^\sigma(O(b)) < 1$. Then $EU_E^\sigma(b) = \sum_{v \in O(a)} P_E^\sigma(v) \cdot u(b, v) < \sum_{v \in O(a) \cap O(b)} P_E^\sigma(v) \cdot u(a, v) + \sum_{v \in O(a) \setminus O(b)} P_E^\sigma(v) \cdot u(a, v) = EU_E^\sigma(a)$. With $K_E := \{v \in \Omega \mid P_E^\sigma(v) > 0\}$ it follows that $b \notin \mathcal{B}(K_E)$, and by (10) that $b \notin \mathcal{B}(A)$. Hence, $b \in \mathcal{B}(A)$ implies $P_E^\sigma(O(b)) = 1$.

Let $A^+ := \bigcap_{a \in \mathcal{B}(A)} O(a)$. We first show that $A^* \cap A^+ \subseteq H$ implies $A \mathrel{+\!>} H$. Let $\hat\sigma \in \mathcal{S}$ be such that $A \in \mathrm{Op}_{\hat\sigma}$. We have to show that $P_E^{\hat\sigma}(H) = 1$. By (15) $P_E^{\hat\sigma}(A^+) = P_E^{\hat\sigma}(\bigcap_{a \in \mathcal{B}(A)} O(a)) = 1$ and by (5) $P_E^{\hat\sigma}(A^*) = 1$; hence $P_E^{\hat\sigma}(A^+ \cap A^*) = 1$, and it follows that $P_E^{\hat\sigma}(H) = 1$.

Next, we show $A \mathrel{+\!>} H$ implies $A^* \cap A^+ \subseteq H$. Suppose that $A^* \cap A^+ \not\subseteq H$. Let $w \in A^* \cap A^+ \setminus H$. From condition $\forall v \in \Omega\, \exists \hat\sigma \in \mathcal{S}\, P_E^{\hat\sigma}(v) = 1$ it follows that there is a support problem $\hat\sigma$ such that $P_E^{\hat\sigma}(w) = 1$. As $w \in A^+$, it follows by (10) that $A \in \mathrm{Op}_{\hat\sigma}$. Due to $A \mathrel{+\!>} H$, it follows that $P_E^{\hat\sigma}(H) = 1$, in contradiction to $w \notin H$.

A^* is the equivalent to the common ground updated with A. In the context of a support problem, we can interpret an answer A as a *recommendation* to choose one of the actions in $\mathcal{B}(A)$. We may say that the recommendation is *felicitous* only if all recommended actions are optimal. Hence, A^+ represents the information that follows from the felicity of the speech act of recommendation which is associated to the answer. It should also be mentioned that $\mathcal{B}(A) = \mathcal{B}(A^*)$ by Definition 4; hence $\bigcap_{a \in \mathcal{B}(A)} O(a) = \bigcap_{a \in \mathcal{B}(A^*)} O(a)$

It is not uninteresting to see that the expert assumption on its own does not guarantee that an utterance has non-trivial implicatures. There are sets $\mathcal{S}$ in which the conditions of Lemma 6 hold but in which answers only implicate what they entail:

Proposition 7 *Let $\mathcal{S}$ be a set of support problems with joint decision problem $\langle(\Omega, P_I), \mathcal{A}, u\rangle$. Assume that for all $X \subseteq \Omega$, $X \neq \emptyset : \exists \sigma \in \mathcal{S}\, K_E^\sigma = X$ and $\exists a \in \mathcal{A}\, O(a) = X$. Then, for all $\sigma \in \mathcal{S}$ with $A \in \mathrm{Op}_\sigma$ it holds $\forall H \subseteq \Omega : A \mathrel{+\!>}_\sigma H \Leftrightarrow A^* \subseteq H$.*

Proof. Condition $\forall X \neq \emptyset\, \exists a \in \mathcal{A}\, O(a) = X$ trivially entails that E is an expert for all $\sigma \in \mathcal{S}$. Condition $\forall X \neq \emptyset\, \exists \sigma \in \mathcal{S}\, K^{\sigma}_{E} = X$ entails the second condition of Lem. 6: $\forall v \in \Omega\, \exists \sigma \in \mathcal{S}\, P^{\sigma}_{E}(v) = 1$. Then, let $A \in \mathrm{Op}_{\sigma}$ and let a^* be such that $O(a^*) = A^*$; as $\mathcal{B}(A) = \mathcal{B}(A^*)$, it follows that $\bigcap\{O(a)\,|\,a \in \mathcal{B}(A)\} = \bigcap\{O(a)\,|\,a \in \mathcal{B}(A^*)\} = O(a^*) = A^*$. Hence, by Lem. 6, $A +\!\!>_{\sigma} H$ iff $A^* \subseteq H$.

This proposition also shows that the conditions of Lemma 6 are less restrictive than they might seem to be.

2.3 Examples

In this section we consider two examples. In both, the answering expert knows the actual state of affairs. This means that we can use Lemma 6 for calculating implicatures. For more examples, we refer to (Benz & v. Rooij, 2007). We start with the classical Out-of-Petrol example (Grice, 1989, p. 31):

(2) I: I am out of petrol.
E: There is a garage round the corner. (G)
+> The garage is open. (H)

We can assume that B's assertion is an answer to the question "*Where can I buy petrol for my car?*" We distinguish four worlds $\{w_1, w_2, w_3, w_4\}$ and two actions $\{\text{go-to-g}, \text{search}\}$. Let d be the place of the garage. Let $G(d)$ mean that d is a petrol station, and $H(d)$ that the place is open. Let the worlds and utilities be defined as shown in the following table:

Ω	$G(d)$	$H(d)$	go-to-d	search
w_1	+	+	1	ε
w_2	+	−	0	ε
w_3	−	+	0	ε
w_4	−	−	0	ε

The answering expert knows that he is in w_1. We assume that P_I and ε are such that $EU_I(\text{go-to-d}|G(d)) > \varepsilon$, i.e. the inquirer thinks that the expected utility of going to that garage is higher than doing a random search in the town. Hence $\mathcal{B}(G(d)) = \{\text{go-to-d}\}$. We see that $O(\text{go-to-d}) = \{w_1\} \subseteq H(d)$. Hence, by Lem. 6, it follows that $G(d) +\!\!> H(d)$.

As second example, we choose an example with several equally useful answers which make the speaker choose different actions:

(3) Somewhere in the streets of Amsterdam...
a) I: Where can I buy an Italian newspaper?
b) E: At the station and at the Palace but nowhere else. (S)
c) E: At the station. (IN(s)) / At the Palace. (IN(p))

IN(d) stand for *d has Italian newspapers*. The answers IN(s), IN(p) and S are equally useful with respect to conveyed information and the inquirer's goals. The answer in b) is called *strongly exhaustive*; it tells us for every location whether we can buy there an Italian newspaper or not. The answers IN(s) and IN(p) are called *mention-some* answers. All answers are optimal, and neither IN(s) implicates that ¬IN(p), nor IN(p) that ¬IN(s). There are two relevant domain objects, the palace p and the station s. We arrive at a model with four worlds and the utilities shown in the following table:

Ω	IN(p)	IN(s)	go-to-p	go-to-s	search
w_1	+	+	1	1	ε
w_2	+	−	1	0	ε
w_3	−	+	0	1	ε
w_4	−	−	0	0	ε

We assume that the probabilities of IN(p) and IN(s) are equal, and that $0 < \varepsilon < 1$. For $d = p, s$, it is $EU_I(\text{go-to-}d|\text{IN}(d)) = 1 > \varepsilon$ and $\mathcal{B}(\text{IN}(d)) = \{\text{go-to-}d\}$. We find $O(\text{go-to-}p) = \{w_1, w_2\} = [\![\text{IN}(p)]\!]$ and $O(\text{go-to-}s) = \{w_1, w_3\} = [\![\text{IN}(s)]\!]$. As $[\![\text{IN}(p)]\!] \not\subseteq [\![\neg\text{IN}(s)]\!]$ and $[\![\text{IN}(s)]\!] \not\subseteq [\![\neg\text{IN}(p)]\!]$, it follows by Lemma 6 that neither IN(p) implicates ¬IN(s), nor IN(s) implicates ¬IN(p).

3 Normal Optimal Answer Models

In this section, we address the question of how to find an OA model given the information we find in a discourse like (1). We will see that normality assumptions play a crucial role in constructing these models. We therefore call these OA models *normal* OA models.

In our introductory example (1), the implicature depends on contextually stated preferences for certain attributes of domain objects. In this section we introduce the *multi-attribute utility theory*[1] (MAUT), a sub-field of applied decision theory which studies decision problems with multiple objectives. We apply it to modelling desires as e.g. expressed by '*I want to buy cheap red wine.*' We will see how the utility function of the OA model can be replaced by a Boolean combination of certain elementary events which describe the success conditions of actions in terms of basic attributes. This allows us to succinctly represent payoff functions by predicates *Good* which take domain objects as their arguments. For example, in case of '*I want to buy cheap red wine*' this predicate is defined by $\mathit{Good}(d) \Leftrightarrow \mathit{Italian\text{-}wine}(d) \wedge \mathit{cheap}(d)$. After having justified this representation, we study an extended example and show how MAUT and a number of *normality* assumptions enter into the construction of OA models. On the basis of this example, we then describe a general procedure for constructing normal OA models.

3.1 Multiple attributes and utilities

In a game or decision theoretic model preferences are represented by a *utility-* or *payoff* function. It assigns to every outcome of the game or decision problem

[1] See (Keeney & Raiffa, 1993), (French, 1988, Ch. 4)

a utility value which represents the overall preferences of a player. In practice, these preferences are often the result of balancing different aspects of the outcome against each other. When buying a new car, the customer may compare different offers according to e.g. price, fuel consumption, insurance group rating, noise level, equipment, warranties, and resell price. Each of these aspects can be evaluated in isolation. They define different *attributes* of the cars. In order to come to a decision, the customer may first evaluate each of the offers according to each attribute, weight the attributes against each other, and then choose according to the result. In practical applications of decision theory, e.g. for modelling the consumer behaviour in retail shopping, the main task is to find a list of relevant attributes $a_1, \ldots, a_n$ and a function U, such that $U(a_1, \ldots, a_n)$ represents the overall payoff function u of an agent. This means, if the agent has to decide between products $d_1, \ldots, d_m$, then $U(a_1(d_1), \ldots, a_n(d_1)), \ldots, U(a_1(d_m), \ldots, a_n(d_m))$ represent his preferences over the outcomes of buying $d_1, \ldots, d_m$. We call the function U a *multi-attribute utility* function. For a sequence of attribute values $a_1(d_j), \ldots, a_n(d_j)$ of a domain object d_j, we write $\mathbf{a}(d_j)$, and for the sequence $a_1, \ldots, a_n$ of attribute-value functions, we write $\mathbf{a}$.

Multi-attribute utility theory investigates the properties of multi-attribute utility functions and their applications; especially it is concerned with the constraints which are imposed on the utility functions by the fact that they depend on an array of attribute values. In general, the values $a_i(d_j)$ can be assumed to be real numbers, but they are not utilities themselves. A higher number does not indicate greater desirability. If a_i measures the number of seats in a car, then an intermediate number like 5 may be optimal. If the number of seats is higher or lower, then the utility diminishes. Attributes may represent (measurable) properties like fuel consumption per 100km or (categorical) properties like '*has air conditioning*'. In the latter case, the attributes have Boolean values and we can set $a_i(d_j) = 1$ if d_j has property a_i, and $a_i(d_j) = 0$ if it doesn't. In the following, we consider only cases in which the attributes represent categorical properties. We use MAUT in particular for modelling attitude reports expressing desires, as e.g.:

(4) **a)** John likes vanilla ice.
b) Peter wants to buy a bottle of cheap red wine.

The sentence (4a) says that John likes ice cream which has the property to be vanilla ice. We cannot directly transform statements in (4) into statements about preferences. Statements about desires have an absolute form. In decision theory, we have to translate such absolute statements into statements about preferences over outcomes of actions. This is our next goal.

Representing the property *of having vanilla taste* by an attribute a with values $0, 1$, and quantifying over the objects d between which J has to choose, we arrive at a formula of the form $\forall d$ (J desires d iff $a(d) = 1$). Doing this also for (4b), we arrive at the following logical forms:

(5) **a)** $\forall d$ (J desires d iff $a(d) = 1$).
b) $\forall d$ (J desires to *verb* d iff $a(d) = 1$).

The formulas in (5) state conditions for when J desires something. They are not yet statements about the preferences expressed by desires. If (4a) is true, then the attribute '*having vanilla taste*' is a relevant attribute of ice cream in situations in which John has to choose between different instances of ice cream. Hence, we explicate the predicate *desire* in (5) as a short form for a statement about preferences over attribute combinations. These preferences over attribute combinations can be represented by a multi-attribute utility function U. In a case like (5a), in which there is only one relevant attribute a, we replace our first approximation $\forall d$ (J desires d iff $a(d) = 1$) by the bi-conditional:

$$\forall d, d' \, [U(a(d)) > U(a(d')) \leftrightarrow a(d) = 1 \wedge a(d') = 0]. \tag{16}$$

Now we have translated the absolute statement '*John desires d iff d has property a*' into a statement about preferences.

If there is only one attribute that influences decisions, then this representation seems to be appropriate. But we may find more complex descriptions of preferences:

(6) **a)** I like vanilla ice and I like strawberry ice.
b) I like vanilla ice but only together with strawberry ice.
c) I like vanilla ice and I like strawberry ice, but I don't like them together.

Every statement describes a complex condition that depends on the attributes of ice cream. These conditions naturally translate into a condition *Good* on the combination of attribute values $\mathbf{a}(d)$ of domain objects d. In general, a multi-attribute utility function U depends on the whole sequence $\mathbf{a}$ of attribute-value functions. Hence, we replace the simple condition $a(d) = 1 \wedge a(d') = 0$ in (16) by a condition involving a predicate *Good*. This leads to a more general logical form for J likes $\mathit{Good}(\mathbf{a}(d))$:

$$\forall d, d' \, [U(\mathbf{a}(d)) > U(\mathbf{a}(d')) \leftrightarrow \mathit{Good}(\mathbf{a}(d)) \wedge \neg \mathit{Good}(\mathbf{a}(d'))]. \tag{17}$$

In (4b) and (5b), there is an action type *act* which is applied to different domain objects d. As there is only one action type, we can describe the preferences over action-argument pairs *act-d* by a condition *Good* over attributes of domain objects. In support problems, the utility measures u are defined as functions which map world-action pairs to real numbers. If the action set is generated by a single action type *act* and a set of domain objects d, then (17) leads to the condition:

$$u(v, \mathit{act}\text{-}d) > u(v, \mathit{act}\text{-}d') \;\; \mathit{iff} \;\; v \models \mathit{Good}(\mathbf{a}(d)) \wedge \neg \mathit{Good}(\mathbf{a}(d')). \tag{18}$$

As payoff functions are unique only up to linear rescaling, we can set without loss of generality:

$$u(v, \mathit{act}\text{-}d) = \begin{cases} 1 & \text{if } v \models \mathit{Good}(\mathbf{a}(d)) \\ 0 & \text{if } v \models \neg \mathit{Good}(\mathbf{a}(d)) \end{cases} \tag{19}$$

It is this special form which we use when constructing normal OA models in Sec. 3.3. It allows us to concentrate on the predicate *Good* when setting up the model.

Finally, we want to line out how our analysis is related to so-called *ceteris paribus* interpretations of preference statements, e.g. (Hanson, 2001). Let's turn to the simple case in which the utilities only depend on a single attribute as e.g. in (16). It may be argued that, in fact, in a statement like (4a), it is meant that John prefers vanilla ice over other flavours of ice cream *all else being equal* (ceteris paribus). Let us assume that U depends on a sequence $a_0, \ldots, a_n$ of attribute-value functions one of which represents the attribute *vanilla taste*. Let's say this is attribute a_0. Then the logical form of a preference statement like (4a) under a ceteris paribus interpretation is:

$$\forall d, d' \, [(a_0(d) = 1 \land a_0(d') = 0 \land \forall i \neq 0 \, a_i(d) = a_i(d')) \rightarrow U(\mathbf{a}(d)) > U(\mathbf{a}(d'))]. \tag{20}$$

In this case, the value of attribute *vanilla flavour* a_0 is only decisive if all other attribute values of the two domain objects are equal. If we only compare two domain object for which all attributes $a_1, \ldots, a_n$ have the same value, then we can disregard them and (20) is equivalent to (16). But, in general, (20) and (16) are not equivalent.

Often, background assumptions which influence an agent's preferences are not explicitly stated. For example, in the situation where a customer has to choose between different cars, he may state that he prefers cars with low fuel consumption. But probably he would be very surprised if the sales attendant thereupon showed him their offer of solar cars. The stated preferences refer to a salient set of relevant alternatives. This may seem to favour the ceteris paribus condition as it asks all alternatives to be equal with respect to all other attributes. But we think this is too strong a condition. For example, if the customer states that he prefers cars with low fuel consumption, then this renders the colour of the car irrelevant. It is possible to decide on the basis of the customer's preference statement that he prefers a green car with low fuel consumption over a blue car with high fuel consumption. This would not be possible under a ceteris paribus interpretation. Hence, we prefer an interpretation of preference statements which assumes that utilities depend on the explicitly stated attributes only *as long as everything else is normal.* The *normality* condition is not made explicit in (16). In what follows, we treat it as an implicit assumption. Which condition is appropriate, the ceteris paribus condition (20) or (16), is an empirical question which only the analysis of examples can answer.

The aim of the next section is to uncover a number of implicit assumptions which we make when interpreting examples similar to (1). We may distinguish two types of normality assumptions. Some explicitly enter into the construction of OA models, some remain implicit in the background. We have just mentioned the assumption that attributes for which no preferences are explicitly stated are disregarded, and it is assumed that the domain objects are *normal* instances with respect to them. This normality assumption is an assumption which we make when setting up a normal OA model, but it is *not* one that is made explicit in the model itself but remains implicit in the background; i.e. we don't find an explicit representation of these disregarded attributes in the model, nor a condition which tell what counts as being normal with respect to them. What this section aims at are the normality assumptions which must be explicitly represented in the OA model in order to account for a) the optimality of answers and b) their implicatures. Finding these normality assumptions is a theory-driven process. Hence, we have to check at the end that the existence of these assumptions is also supported by examples.

3.2 The analysis of an example

Preferences may depend on attributes in different ways, as indicated in (17). In the following example, the decision depends on several attributes which must all be satisfied to make the outcome optimal:

(7) J: I am looking for a house to rent in a quiet and safe neighbourhood, close to the city centre, with a small garden and a garage.
E: I have an offer for you in Dorothy Avenue.

The answer (7) clearly implicates that the offered house is located in a quiet and safe neighbourhood close to the city centre and has a small garden and a garage. The same pattern occurs if only one but not all of the attributes must be satisfied:

(8) **a)** A: I want to see a classical Beijing opera tonight or Chinese acrobatics, but I don't want to go to one of these modern tea houses which mix both things. What can I do tonight?
B: You can go to the Lantern Tea House!

b) John loves to dance to Salsa music and he loves to dance to Hip Hop, but he can't stand it if a club mixes both styles.
J: I want to dance tonight. Is the Music in Roter Salon ok?
E: Tonight they play Hip Hop at the Roter Salon.

In (8a), the answer implicates that the Lantern Tea House shows either a classical Beijing opera or a Chinese acrobatics performance but not both, and in (8b) that there is only Hip Hop at the Roter Salon.

We consider the last example (8b) in more detail. In accordance with our assumptions in Section 2 we assume that the answering expert is fully cooperative and knows all about the inquirers expectations. In addition we assume that he knows that there plays Hip Hop at the Roter Salon and not Salsa. The first two lines of (8b) introduce the relevant attributes and the properties of the inquirer's utility function. There are two attributes. Let d range over dance locations; we write:

$H(d)$: There is Hip Hop at d; $S(d)$: There is Salsa at d.

What has to be explained by an OA model is: a) why is $H(d)$ an optimal answer, and b) why does $H(d)$ implicate that not $S(d)$? What we now want to find out is which assumptions we have to make in addition to what is explicitly stated in (8b) in order to answer these questions. We assume that these additional assumptions are automatically accommodated.

The first task in the application of the OA approach is the construction of the support problem $\langle \Omega, P_E, P_I, \mathcal{A}, u\rangle$ which models the answering situation after J's question '*Is the Music in Roter Salon ok?*' From the first sentence of J's utterance, we know that this question serves to solve his decision problem of where to go for dancing. Hence, we can conclude that the action set $\mathcal{A}$ consists of all acts *going-to-d* in which d ranges over dance locations. In addition, we can assume that there is an alternative act *stay-home*.

The background provided in the first sentence tells us how the utility function u is defined. We can translate the sentence into a property of dance locations: $\mathit{Good}(d)$ *iff* $(H(d) \vee S(d)) \wedge \neg(H(d) \wedge S(d))$. The arguments of u are world-action pairs (v, a). Obviously, the outcome of performing act *going-to-d* in world v is desired iff $v \models \mathit{Good}(d)$, hence, we can assume that:

$$u(v, \mathit{going\text{-}to\text{-}d}) = \begin{cases} 1 \Leftrightarrow v \models \mathit{Good}(d); \\ 0 \Leftrightarrow v \models \neg\mathit{Good}(d). \end{cases} \tag{21}$$

There are three assumptions that enter here. First, we have to assume that $H(d)$ and $S(d)$ are the only attributes that count, i.e. there are no hidden objectives which are relevant for finding an optimal answer. Second, the objectives $H(d)$ and $S(d)$ are of equal weight, i.e. J does not prefer e.g. $H(d) \wedge \neg S(d)$ over $S(d) \wedge \neg H(d)$. Third, (21) also supposes that $H(d) \wedge S(d)$ is equally dis-preferred to $\neg H(d) \wedge \neg S(d)$.

This leaves us with the problem of finding $\langle \Omega, P_E, P_I\rangle$, which is neither given by the background, nor by the question itself. We first consider the probability distributions P_E and P_I.

We first turn to P_E. It is clear that E's answer can only implicate that there is no Salsa playing at Roter Salon if it is common knowledge that E knows which music they are playing at Roter Salon. Hence, if r is the Roter Salon, then we must assume that $P_E(\mathit{Good}(r)) = 1$ or $P_E(\mathit{Good}(r)) = 0$.

We now turn to P_I. In (8b), nothing is stated about the inquirer's prior expectations about the music playing at different locations. Hence, we may assume that the inquirer has no reason to expect that Salsa playing at a location d is more probable than Hip Hop playing at a location e. This leads to the assumption that for $X \in \{H, S\}$: $\exists \alpha \in (0,1)\, \forall d\, P_I(X(d)) = \alpha$.

We can also see that in (8b) nothing is said about the dependencies between events. Hence, we can assume that Salsa playing at one place is independent of Hip Hop playing at another place. This leads to the assumption that for all places d, e, $d \neq e$, the events $H(d)$, $S(d)$, $H(e)$ and $S(e)$ are probabilistically independent, which means that for all non-empty subsets $\mathcal{E} \subseteq \{H(d), S(d), H(e), S(e)\}$: $P_I\left(\bigcap \mathcal{E}\right) = \prod_{X \in \mathcal{E}} P_I(X)$.

Definition (21) entails that $EU_I(\mathit{going\text{-}to\text{-}d}) = P_I(\mathit{Good}(d))$. If α is the probability of the elementary events $H(d)$ and $S(d)$, then, by independence, it follows that $P_I(\mathit{Good}(d)) = 2\alpha(1-\alpha)$. Furthermore, for $d \neq e$ and $X \in \{H, S, \mathit{Good}\}$ it is $P_I(\mathit{Good}(d) \wedge X(e)) = P_I(\mathit{Good}(d)) \times P_I(X(e))$.

Furthermore, we have to make an assumption about the relation between the expected utilities of *stay-home* and *going-to-d* before and after learning that $S(d)$ or $H(d)$. Let $X \in \{H, S\}$, then we assume:

$$EU_I(\textit{going-to-d}) < EU_I(\textit{stay-home}) < EU_I(\textit{going-to-d}\,|X(d)). \quad (22)$$

This means, the inquirer believes that it is more probable that a place plays only Salsa or only Hip Hop. If we do not make this assumption, then the addressee might still decide to stay at home after learning that $H(r)$, in which case $H(r)$ would not be an optimal answer as it fails to induce the addressee to choose an optimal action. Hence, the assumption is necessary for explaining the optimality of $H(r)$. We will say more about neutral acts like *stay-home* later on.

We now show that these assumptions indeed entail the implicatures observed in Example (8b). Let r denote the *Roter Salon*. Then, after learning that $H(r)$, we find for all other places d:

$$P_I(\mathit{Good}(d)|H(r)) = P_I(\mathit{Good}(d)) < P_I(\mathit{Good}(r)|H(r)). \quad (23)$$

Assumption (22) together with proposition (23) imply that learning $H(r)$ will induce the hearer to go to the Roter Salon. As $EU_E(H(r)) = 1$ and as there is no possible answer A such that $EU_E(A) > 1$, it follows that $H(r)$ is optimal. The following table shows a model for (8b). As we assumed that the speaker knows that only Hip Hop is played at Roter Salon, he knows that the actual world is w_2:

(9)

Ω	$H(r)$	$S(r)$	*stay-home*	$\mathit{Good}(r)$	$H(r) \in \mathrm{Op}_{w_j}$
w_1	1	1	ε	0	no
w_2	1	0	ε	1	yes
w_3	0	1	ε	1	no
w_4	0	0	ε	0	no

Let, as usual, $[\![\varphi]\!] = \{w \in \Omega \mid w \models \varphi\}$. Then $O(\textit{going-to-r}) = [\![\mathit{Good}(r)]\!] = \{w_2, w_3\}$. As $[\![H(r)]\!]^* = [\![H(r)]\!] = \{w_1, w_2\}$, we find $[\![H(r)]\!]^* \cap O(\textit{going-to-r}) = \{w_2\} = H(r) \setminus S(r)$. From Lemma 6, it follows that $H(r)$ implicates that only Hip Hop is played at Roter Salon. This proves the claim.

Finally, this leads to the last parameter which was left unexplained, the set of possible worlds Ω. What counts for I's decision are only the truth values of $H(r)$ and $S(r)$. Hence, we can identify the set of possible worlds Ω with the set of all truth value assignments to the formulas $H(r)$ and $S(r)$.

There are two obvious questions to which this example gives rise: First, are the assumptions also necessary for deriving the implicature, and second, is there a general structure which we can extract from this example and apply to the interpretation of similar examples? We start with the second question.

3.3 The construction of normal models

In this section we present a procedure for constructing a *normal* OA model. The construction starts with a given set of domain objects and a sequence of attributes.

In (8b), the first sentence introduces a set of two attributes which we represented before by the formulas $H(x)$ and $S(x)$. John's question introduced the domain object *Roter Salon*. With these two parameters, we can construct a suitable set of possible worlds. Let D be a set of domain objects. We denote attribute-value functions for a fixed sequence of n attributes by $(a_i)_{i=1,\dots,n}$. We assume that all attributes range over values in $\{0,1\}$. Then we set:

$$\Omega := \{(a_i)_{i=1,\dots,n} \mid a_i : D \to \{0,1\}\}. \tag{24}$$

This means, Ω is the set of all attribute-value functions which map the objects in D to their respective values in $\{0,1\}$. In this section, we write $\mathbf{a}$ for the elements of Ω. If $|D| = m$, then $|\Omega| = 2^{mn}$. In (8b), it is $D = \{r\}$ and $n = 2$, hence, $|\Omega| = 4$.

The assumptions that enter into this construction can be summarised as follows:

I Assumptions about Objectives:
1. Only attributes of elements of D count.
2. Completeness of objectives: $(a_i)_{i=1,\dots,n}$ represents all objectives.

We add a strong default assumption about the speaker's knowledge:

II Expert assumption: $\exists \mathbf{a} \in \Omega\ P_E(\mathbf{a}) = 1$.

In analogy to the construction of possible worlds from attribute-value functions in (24), we define the probabilities of these worlds in terms of the probabilities of *elementary events* which are constructed from the same attribute-value functions. For each attribute a_i, we can introduce a prime formula A_i such that $\mathbf{a} \models A_i(d)$ *iff* $a_i(d) = 1$. We then define $[\![\varphi]\!] = \{\mathbf{a} \in \Omega \mid \mathbf{a} \models \varphi\}$, and set $E(n, D) := \{[\![A_i(d)]\!] \mid d \in D, 1 \le i \le n\}$. $E(n, D)$ is a set of elementary events. For example, in (8b), it is the set $\{H(r), S(r)\}$. If the reader of an example has no further information about the elementary events, he has no reason to expect that one event is more probable than the other. Hence, by the *principle of insufficient reason*, he should treat them as equally probable and independent of each other. This is captured by the following conditions:

III Laplacian Assumptions:
1. Equal undecidedness:

$$\forall X, Y \in E(n, D) : P_I(X) = P_I(Y). \tag{25}$$

2. Probabilistic independence:

$$\forall \mathcal{E} \subseteq E(n, D) : P_I\left(\bigcap \mathcal{E}\right) = \prod_{X \in \mathcal{E}} P_I(X). \tag{26}$$

If $P_I(A_i(d))$ is known for all i, d, then the Laplacian assumptions uniquely define P_I. To see this, let $\alpha := P_I(A_i(d))$ and $n_v := |\{(i,d) \mid v \models A_i(d)\}|$. As before, let $m = |D|$ and let n be the number of attributes. Then, $P_I(v) = \alpha^{n_v} \times (1-\alpha)^{mn-n_v}$. In particular, if $\alpha = \frac{1}{2}$, i.e. if the addressee is equally undecided with regard to the truth of $A_i(d)$ and $\neg A_i(d)$, we find $P_I(v) = 2^{-mn}$.

In (8b), the first sentence not only introduced the relevant attributes but also stated a constraint on the multi-attribute utility function U. The sentence stated that dancing at a place d is desirable iff $H(d) \vee S(d)$ but not $H(d) \wedge S(d)$. Let us abbreviate this condition by $\mathit{Good}(d)$. This condition leaves the multi-attribute utility function U highly underspecified, but it is natural to interpret the statement such that the utility of dancing at d is equal in all worlds in which $\mathit{Good}(d)$ holds, and equal in all worlds in which $\neg\mathit{Good}(d)$ holds. This motivates the following normality assumption about utilities:

IV Equal weight of objectives: All values which can be treated as equal are treated as equal.

The last parameter that is needed to set up a normal decision problem for an example is the set of actions. This set is closely connected to the set of domain objects D. Looking again at (8b), we see that the domain objects, the clubs that play dance music, are an argument, in this case the *goal*, of the actions between which the inquirer has to choose. If there is only one type of action, then we can identify the set of hearer's actions with a set $\{\mathit{act}\text{-}d \mid d \in D\}$. A multi-attribute utility function U represents preferences over *states of the world*. These states are the outcomes of actions between which an agent can choose. With U as before, the preferences over world-action pairs must satisfy the following condition:

$$u(\mathbf{a}, \mathit{act}\text{-}d) = U(a_1(d), \ldots, a_n(d)) \tag{27}$$

As argued for in Section 3.1, this leads to a representation of desirability by a payoff function u which is defined over world-action pairs as in (19) using the predicate *Good*.

In (8b), we also introduced a *neutral* action, *stay-home*, or doing nothing. In the Out-of-Petrol example, (2), the neutral act was a random search in town. That we need such assumptions can be best seen from (8b). If the action set only consisted of the acts of going to places d, then, if there is only one place, the inquirer has no choice and any answer would be optimal. Hence, the assumption of a default action is partly a technical device that allows us to reduce the set of possible worlds and the set of hearer's actions to those explicitly stated in the examples. It is clearly desirable to have a uniform characterisation of the default act. In the Out-of-Petrol and the Hip-Hop examples, we could unify their characterisations by assuming that they consist of a random sequence of acts *act-d* which are followed by the act of doing nothing if none of them is successful. Alternatively, we could think of the default act as a lottery over the acts *act-d* followed by the act of doing nothing if *act-d* fails. This depends on whether or not it is reasonable to assume that the acts *act-d* can be performed in sequence. Hence, in general, we can assume that the neutral act consists of a random sequence of acts from $\{act\text{-}d \mid d \in D\}$ followed by an act of doing nothing, including the special case in which the random sequence has length zero. In addition, we have to assume that the length of the sequence causes some costs which are implicitly represented in the payoff function u. The Hip-Hop example as represented in (9) is an example in which the sequence has length zero. This assumption was necessary as D contained only one element r; hence, if the random sequence had length longer than zero, then the first act of the *random* sequence could only be the act of going to r, which means that there is only one act to choose for the hearer, namely going to r. What we need, in any case, is an action set $\mathcal{A}$ which consists of at least two actions with distinct outcomes for which a property similar to (22) can be proven. Furthermore, as it is a default act, it must be the alternative which the hearer would choose if he had to choose before learning anything from the speaker. Although we can't claim to fully understand the nature of the normality assumption which is involved here, we make the following tentative assumptions about the existence of a neutral act l:

V Neutral alternative: i) $l \in \mathcal{B}(\Omega)$ and ii) $\mathcal{B}(K_E) = \{l\} \vee l \notin \mathcal{B}(K_E)$ with $K_E = \{v \in \Omega \mid P_E(v) > 0\}$.

The first condition says that l is a best choice if the hearer doesn't learn any new information. It entails $EU_I(l) > \max_d EU_I(act\text{-}d)$. The second condition means that if the answering expert can recommend any act different from l, then the neutral act l is no longer an optimal choice.

Together, these default rules define a unique support problem σ for each sequence $\langle (a_i)_{i=1}^n, D, U, \alpha, l\rangle$ in which $(a_i)_{i=1}^n$ is a sequence of attributes, D a set of domain objects, U a multi-attribute utility function, α the probability of the elementary events $A_i(d)$, and l the neutral act. We can even simplify this description as only the *numbers* n of attributes and m of domain objects count. Furthermore, in all those cases in which U only ranges over values in $\{0, 1\}$, we can characterise U by a formula $Good(d)$ which only contains formulas of the form $A_i(d)$ with $d = 1, \dots, m$. Finally, if there is no reason to assume otherwise, $\alpha = \frac{1}{2}$. Then, σ only depends on a quadruple $\langle n, m, Good(\,.\,), l\rangle$.

A full justification of these default rules can only be given by examples. Table 2 shows the combined table for examples (8a) and (8b). In both examples, there is one domain object d and two attributes. The speaker recommends an action, *going-to-d*. This recommendation is optimal in world w_i if $w_i \models Good(d)$. This is the case in w_2 and w_3. Hence, this recommendation implicates that the domain object d has either attribute A_1 or A_2, but not both. From this, we can see that the assertion of $A_1(d)$ or $A_2(d)$ implicates that the domain object has only the asserted attribute.

In Table 3, we show the model of (7). Again, there is only one domain object but six attributes. By the rule about equal weights of objectives, $Good(d)$ only holds if it has all attributes $A_1, \dots, A_6$. The speaker recommends an action *act-d*, hence, we have to find out which worlds are elements of $O(act\text{-}d)$. This can only be the world in which d satisfies all attributes. Hence, the recommendation implicates that d has all of the desired attributes. A possible objection may

Ω	$A_1(d)$	$A_2(d)$	$Good(d)$	$A_1(d) \in \text{Op}_{w_i}$
w_1	1	1	0	no
w_2	1	0	1	yes
w_3	0	1	1	no
w_4	0	0	0	no

Table 2. Joint table for examples (8a) and (8b)

arise with respect to the number of domain objects. Whereas in (8b) the inquirer explicitly asks about one object, no such restriction exists in (8a). Hence, one would expect an underspecified number of domain objects. But the representation of these objects is unnecessary due to the existence of the neural act l.

Ω	$A_1(d)$	...	$A_5(d)$	$A_6(d)$	$Good(d)$
w_1	1	...	1	1	1
w_2	1	...	1	0	0
...	...		...	...	...

Table 3. The table for Examples (7)

3.4 Violations of the normality assumptions

In this section we look at the effects of violations of the normality assumptions. We show that their violations also lead to different implicatures. We present a series of variations of (8b) which violate one of the rules. As we assume that the rules hold by default, only an explicit statement of their invalidity or general world knowledge can violate them.

We start with the expert assumption **II**. If it is known that the answering person has only partial knowledge about the music playing at Roter salon, the implicature from $H(r)$ to $\neg S(r)$ may be suspended:

(10) E: I know that Hip Hop is played at the Roter Salon tonight. But I don't know whether they mix it with other music.

Still, this answer may induce the inquirer to go to the Roter Salon, but the implicature that there is no Salsa is clearly cancelled.

The normality assumptions **I** and **IV** are a kind of exhaustification principles. Hence, it is not possible to violate them directly by stating new objectives or more details of the utility function. Hence, we assume that violating facts belong to unstated background knowledge. For example, assume that in (8a) it is common knowledge that the inquirer only wants to go out if he finds a place close by. Then, this additional objective has the effect that B's answer "You can go to the Lantern Tea House" implicates that the Lantern Tea House is close by.

Let us assume that, in Example (8b), it is common knowledge that the inquirer's preferences for Salsa far exceed that for Hip Hop. If he asks "I want to dance tonight. Where can I go to?", then the answers "Tonight they play Hip Hop at the Roter Salon" has the additional implicature that there is no place where they only play Salsa.

Violations of **III** are probably most interesting. In the following example, we find an explicitly stated dependency between playing Hip Hop and playing Salsa:

(11) The Roter Salon and the Grüner Salon share two DJs. On of them only plays Salsa, the other one mainly plays Hip Hop but mixes into it some Salsa. There are only these two Djs, and if one of them is at the Roter Salon, then the other one is at the Grüner Salon. John loves to dance to Salsa music and he loves to dance to Hip Hop but he can't stand it if a club mixes both styles.
J: I want to dance tonight. Is the Music in Roter Salon ok?
E: Tonight they play Hip Hop at the Roter Salon.

Let gr denote the Grüner Salon. The dependencies imply that

$$P_I(Good(r)|H(r)) = 0 \text{ and } P_I(Good(gr)|H(r)) = 1,$$

Hence, after hearing that $H(r)$, the inquirer can infer that going to the Grüner Salon is the optimal choice. This implies that the answer $H(r)$ is also optimal with $EU_E(H(r)) = 1$. Indeed, in this context $H(r)$ implies that the music at Roter Salon will be bad.

The last normality assumption, **V**, is probably the most problematic one. It is also difficult to test. That is mainly because it serves a technical purpose. Besides being a means for simplifying the models, it likewise prevents some unwanted predictions of optimality of answers. Assume that in the Out-of-Petrol example (2), there is a single road which the inquirer can either follow to the left or to the right. To the left, there are fields, bushes, and trees. To the right, there is the beginning of a town. In this scenario, turning to the right is probably more promising than turning to the left. But this entails that saying nothing is already an optimal answer if the answering expert knows that to the right there is a petrol station. This possibility is ruled out when assuming that there exists a neutral act with the properties stated in **V**.

4 A classification of support problems and implicatures of simple answers

The aim of this section is to show how to use the normal optimal answer approach for deriving systematic classifications of support problems representing dialogue situations and calculating the associated lists of implicatures for simple answers. In Section 3.3, we saw that the default rules define a unique support problem for each quintuple $\langle (a_i)_{i=1}^n, D, U, \alpha, l \rangle$ in which $(a_i)_{i=1}^n$ is a sequence of attributes, D a set of domain objects, U a multi-attribute utility function, α the probability of the elementary events $A_i(d)$, and l the neutral act. In this section, we only consider cases in which U can be characterise by a formula $Good(d)$ which is in turn a Boolean combination of formulas $A_i(d)$ representing the multiple basic attributes. In addition, we assume that the probability of all elementary events equals $\frac{1}{2}$. We have seen that these restrictions imply that the normal support problem describing the dialogue situation is uniquely characterised by a quadruple $\langle (a_i)_{i=1}^n, D, Good(\,.\,), l\rangle$. In this section, we consider two types of dialogue situations. First, the type of situations with one domain object and two attributes; and second, the type of situations with two objects and one attribute. The first type of dialogue situations includes the Out-of-Petrol example (2), the second the Italian newspaper example (3). It should become clear how to extend these classifications to the general case, i.e. to the cases with an arbitry number of domain objects and attributes.

In examples (8a) and (8b), we find one relevant domain object and two attributes which have an influence on the decision of the addressee. The model was shown in Table 2. From this table, we can derive all normal models for situations with one domain object and two attributes if we systematically list all possible utility functions which can be described by a Boolean predicate *Good*. The following table shows the definition of the predicate *Good* in examples (8a) and (8b):

(12)

Ω	A	B	*Good*
w_1	1	1	0
w_2	1	0	1
w_3	0	1	1
w_4	0	0	0

In case of example (8b), attribute A denotes the playing of Hip Hop, and B the playing of Salsa at the Roter Salon. *going-to-r* is *good* just in case the Roter Salon has one but not both attributes. The first column of Table 4 shows all possible predicates *Good* which can be defined from two attributes A and B. The four numbers listed in each row tell us whether *Good* is true in the worlds $w_1, \ldots, w_4$ which are defined in (12). Hence, for example, in the fifth row, the entry 1 0 1 1 means that choosing the only domain object is *good* in worlds w_1, w_3, w_4 and inferior to the neutral act in world w_2. The other columns in Table 4 show entries for the four answers A, B, $\neg A$, and $\neg B$. Each row has four entries for each answer. They again relate to the four worlds $w_1, \ldots, w_4$. A '✓' means that the answer makes the hearer choose an act which is optimal in the world to which the ✓ entry belongs; a '$-$' means that the answer will lead the hearer to choose a sub-optimal act; and a '$\cdot$' means that the answer is not admissible. Hence, in the fifth row, the entry $-$, ✓, $\cdot$, $\cdot$ for answer A says that A induces the hearer to choose the inferior act in w_1, it induces him to choose the optimal act in world w_2, and it is not admissible in w_3 and w_4. In this row *Good* is defined such that the neutral act is preferred in w_2, and dispreferred in w_1; hence, the '$-$' and '✓' entries in the first and second position for answer A both say that A induces the hearer to choose the neutral act. In order to see that

Good				A				B				$\neg A$				$\neg B$			
1	1	1	1	✓	✓	·	·	✓	·	✓	·	·	·	✓	✓	·	✓	·	✓
1	1	1	0	✓	✓	·	·	✓	·	✓	·	·	·	–	✓	·	–	·	✓
1	1	0	1	✓	✓	·	·	–	·	✓	·	·	·	✓	–	·	✓	·	✓
1	1	0	0	✓	✓	·	·	–	·	✓	·	·	·	✓	✓	·	–	·	✓
1	0	1	1	–	✓	·	·	✓	·	✓	·	·	·	✓	✓	·	✓	·	–
1	0	1	0	–	✓	·	·	✓	·	✓	·	·	·	–	✓	·	✓	·	✓
1	0	0	1	–	✓	·	·	–	·	✓	·	·	·	✓	–	·	✓	·	–
1	0	0	0	✓	–	·	·	✓	·	–	·	·	·	✓	✓	·	✓	·	✓
0	1	1	1	✓	–	·	·	✓	·	–	·	·	·	✓	✓	·	✓	·	✓
0	1	1	0	✓	–	·	·	✓	·	–	·	·	·	–	✓	·	–	·	✓
0	1	0	1	✓	–	·	·	✓	·	✓	·	·	·	✓	–	·	✓	·	✓
0	1	0	0	–	✓	·	·	✓	·	✓	·	·	·	✓	✓	·	✓	·	–
0	0	1	1	✓	✓	·	·	✓	·	–	·	·	·	✓	✓	·	✓	·	–
0	0	1	0	✓	✓	·	·	–	·	✓	·	·	·	✓	–	·	✓	·	✓
0	0	0	1	✓	✓	·	·	✓	·	✓	·	·	·	–	✓	·	–	·	✓
0	0	0	0	✓	✓	·	·	✓	·	✓	·	·	·	✓	✓	·	✓	·	✓

Table 4. A complete list of all *Good* predicates for the Out-of-Petrol examples.

A indeed recommends the neutral act, we must compare the expected utilities of the different acts. For each row, it is assumed that the expected utility of the neutral act is minimally higher than the expected utility of choosing the domain object. Hence, if learning A increases the probability of *Good*, then this is interpreted as a recommendation to choose the domain object, otherwise it is interpreted as a recommendation to choose the neutral act. In the fifth row, learning A decreases the probability of A, hence it recommends the neutral act.

As second example for the systematic application of normal models, we consider the case of two domain objects with one attribute. This class of examples includes the Italian newspaper example (3). The table in (13) shows the four possible worlds which can be defined by two objects a, b and one attribute A:

(13)

Ω	$A(a)$	$A(b)$
w_1	1	1
w_2	1	0
w_3	0	1
w_4	0	0

As the predicate $Good(d)$ must be a Boolean combination of the prime formula $A(d)$, there are four possible definitions. They are listed in the first column of the next table:

(14)

$Good(d)$	$Good(a)$	$Good(b)$
$A(d) \vee \neg A(d)$	1 1 1 1	1 1 1 1
$A(d)$	1 1 0 0	1 0 1 0
$\neg A(d)$	0 0 1 1	0 1 0 1
$A(d) \wedge \neg A(d)$	0 0 0 0	0 0 0 0

The four entries for the formulas $Good(a)$ and $Good(b)$ show whether the respective formulas are true or not in the worlds $w_1, \dots, w_4$ defined in (13).

We again assume that the expected utility of the neutral act is minimally higher than the expected utility of choosing one of the domain objects a, b. Hence, if learning an answer increases the probability of $Good(d)$, then this is interpreted as a recommendation to choose a domain object. The table in (15) shows for each definition of $Good(d)$ and each of the worlds $w_1, \dots, w_4$ whether the simple answers $A(a)$, $A(b)$, $\neg A(a)$, and $\neg A(b)$ are optimal:

(15)

$A(a)$	$A(b)$	$\neg A(a)$	$\neg A(b)$
✓ ✓ · ·	✓ · ✓ ·	· · ✓ ✓	· ✓ · ✓
✓ ✓ · ·	✓ · ✓ ·	· · − ✓	· − · ✓
✓ − · ·	✓ · − ·	· · ✓ ✓	· ✓ · ✓
✓ ✓ · ·	✓ · ✓ ·	· · ✓ ✓	· ✓ · ✓

Each row corresponds to one of the definitions of $Good(d)$ shown in (14), and the four entries for each answer to the worlds $w_1, \dots, w_4$. Hence, the entry ✓, −, ·, · in the third row below $A(a)$ says that if $Good(d)$ is equivalent to $\neg A(d)$, then answer $A(a)$ induces the hearer to choose the optimal action in world w_1 but misleads him in w_2. This can be calculated as follows: as learning $A(a)$ does not increase the probability of $Good(b)$, the neutral act still has the highest expected utility; hence, answer $A(a)$ must be interpreted as a recommendation to choose the neutral act, which is optimal in w_1 but not in w_2.

Among other reasons, the neutral act was introduced for representing the effect of unmentioned alternative domain objects which are in the choice set of the hearer. If the neutral act consists in first trying two objects a and b in a random order, and then, if unsuccessful, doing nothing, then an answer which discourages the choice of one domain object d will be equivalent to a recommendation of the other object. This changes the table in (15). The result is shown in (16):

(16)

$A(a)$	$A(b)$	$\neg A(a)$	$\neg A(b)$
✓ ✓ · ·	✓ · ✓ ·	· · ✓ ✓	· ✓ · ✓
✓ ✓ · ·	✓ · ✓ ·	· · ✓ −	· ✓ · −
− ✓ · ·	− · ✓ ·	· · ✓ ✓	· ✓ · ✓
✓ ✓ · ·	✓ · ✓ ·	· · ✓ ✓	· ✓ · ✓

As an example, we consider the following variation of the Italian newspaper example (3):

(17) I: I heard that I can buy Italian newspapers at the station or at the Palace. Do you know where I should go to?
E: The Palace doesn't sell Italian newspapers.

If we identify the Palace with domain object a, then looking into the second row under $\neg A(a)$ in (16) shows us that answer $\neg A(a)$ can only be successful in world w_3. Hence, the hearer can infer that $A(b)$, i.e., that he can buy an Italian newspaper at the station.

These two examples, the situations with one domain object and two attributes, and the situations with two objects and one attribute, show the potential of normal optimal answer models for categorising dialogue situations and systematically deriving the associated lists of optimal answers. At the same time, these examples also point to some limitations of the model. Let us consider Example (17) again. The negated answer $\neg A(a)$ is generally classified as a partial answer which implicates that the speaker does not have full domain knowledge. Hence, the answer leads to a suspension of the normality assumption that the speaker is a domain expert. In order to arrive at the predicted implicature, it has to be common knowledge that there can be no doubt that the speaker is an expert. A similar problem arises with the Out-of-Petrol examples in Table 4. We consider the following example:

(18) **a)** I: I want to have a house with either both a garden and a balcony, or with neither a garden nor a balcony.
b) I: I want to have a house with both a garden (A) and a balcony (B).
c) E: The house in Shakespeare Avenue has a balcony (A).

For this example, the entries in the seventh and eighth row below A in Table 4 tell us that w_2 must be the actual world in (18a), and w_1 in (18b). Hence, answer A implicates that the house in Shakespeare Avenue does not have a garden in (18a), and does have a garden in (18b). Probably, the most natural continuation would in both cases be a clarification request of the form: '*And does it also have a garden?*' The effect of the possibility of clarification requests has not been accounted for in our normal optimal answer model. Hence, we can detect at least two shortcoming of the models proposed in this paper: they cannot deal with the possible suspension of normality assumptions, and they cannot account for the effects of clarification requests. A first step towards a solution to the first problem was taken in (Benz, 2009). The effect of clarification requests was studied in (Benz, 2008). In both cases, a fully satisfying solution has to wait for future research.

5 Conclusion

In the introduction, we characterised our paper as a preparatory study for a theory of relevance implicatures in discourse. What it achieves is an isolation of the crucial parameters and a list of normality assumptions which are needed for the construction of normal optimal answer models. That this is far from a discourse interpretation theory for implicatures needs no further explanation. What our study tells us is which parameters we have to look for when interpreting text in order to obtain its implicatures. What is most clearly needed is a theory of the interaction of discourse structure, as defined by rhetorical relations, and conversational implicatures which are defined by background decision problems. We must leave this to future research.

That multi-attribute utility theory should play a major role in the analysis of relevance implicatures is not too surprising in view of the theory's importance to applied decision theory. There, the elicitation of the decision maker's preferences over attributes is a central step in setting up the model of a decision problem. And in view of the importance of decision making in everyday life, it is also not too surprising that language users have developed the means for efficiently communicating these preferences.

Attribute-value functions play two important roles in our model. First, as they represent the attributes to which the agents react. This means, they tell us what the relevant properties of the epistemically possible worlds are. This led us to identify possible worlds with the attribute-value functions which we find by the multi-attribute utility analysis of a given discourse. The second important application of attribute utility functions concerned the representations of desires. We assumed that a sentence like "*John likes vanilla ice*" is adding a new attribue and setting the value of John's utility function to a default value. Here again, our paper is only clearing some ground where a comprehensive theory must be developed in the future.

The last section showed how to derive systematic classifications of dialogue situations based on the normal optimal answer approach. We also showed how to calculate the associated lists of optimal answers. This section reveals some limitations of our approach. In particular, there is no account for the effect of clarification request, or for the possibility of suspending a normality assumption. A satisfying solution, again, has to wait for future research. Although there are clear limits to the model, the underlying normality assumptions, which are the main concern of this paper, are not affected by them. These assumptions must play an important role in any discourse interpretation theory which not only accounts for semantic and rhetoric information but also for Gricean implicatures.

Bibliography

A. Benz (2006): *Utility and Relevance of Answers;* in: A. Benz, G. Jäger, R. v. Rooij (eds.): *Game Theory and Pragmatics*, Palgrave, Mcmillan.

A. Benz (2008): *Implicatures of Irrelevant Answers and the Principle of Optimal Completion;* In: P. Bosch et al. (eds.): 7th International Tbilisi Symposium on Logic, Language, and Computation, TbiLLC 2007, Tbilisi, Georgia, October 2007, Revised Selected Papers. Springer, Lecture Notes in Artificial Intelligence, Berlin, pp. 95–109.

A. Benz (2009): *On the Suspension of Implicatures;* Proceedings of the 10th Symposium on Logic and Language, LoLa10, Budapest, pp. 73–79.

A. Benz, R. v. Rooij (2007): *Optimal assertions and what they implicate: a uniform game theoretic approach;* In *Topoi - an International Review of Philosophy* 27(1), 2007, pp. 63-78 .

S. French (1988): *Decision Theory — an introduction to the mathematics of rationality;* Ellis Harwood, Chichester.

H.P. Grice (1989): *Logic and Conversation;* in: H.P. Grice: *Studies in the Way of Words*, Harvard University Press, Harvard, pp. 22-40.

J. Groenendijk, M. Stockhof (1991): *Dynamic predicate Logic;* Linguistics & Philosophy 14, pp. 39—100.

S.O. Hanson (2001): *The Structure of Values and Norms;* Cambridge University Press, Cambridge.

R.L. Keeney, H. Raiffa (1993): *Decisions with Multiple Objectives —' Preferences and Value Tradeoffs;* Cambridge University Press, Cambridge, (first published by John Wiley & Sons, 1976).

D. Lewis (2002): *Convention;* first published by Harvard University Press 1969; reissued by Blackwell Publishers Ltd., Oxford.