Implicatures of Irrelevant Answers and the Principle of Optimal Completion

Anton Benz

Center for General Linguistics, Schützenstrasse 18, 10117 Berlin, Germany

Abstract. In this paper, we present a game theoretic account of a subclass of '*relevance*' implicatures arising from *irrelevant* answers. We show that these phenomena can be explained if we assume that interlocutors agree on production and interpretation strategies that are robust against small 'trembles' in the speaker's production strategy. In this context, we argue for a new pragmatic principle which we call the *principle of optimal completion*. We als show that our model provides a parallel account of scalar implicatures which removes some limitations of previous accounts.

1 Introduction

The pragmatic appropriateness of answers and their implicatures in decision contexts has been a major topic in the field of game theoretic pragmatics, see e.g. [6, 8, 9]. In [3], a uniform account was given for scalar and relevance implicatures arising in decision contexts. This account was based on the *optimal–answer* (OA) model [1]. In this paper, we address some open problems which arise in connection with apparently irrelevant answers. This will lead to a major revision and improvement of the OA framework. The crucial examples are derived from the classical *Out-of-Petrol* Example [4]:

(1) H is standing by an obviously immobilized car and is approached by S, after which the following exchange takes place:
H: I am out of petrol.
S: There is a garage round the corner. (G)
+> The garage is open. (I)

In the OA approach, the implicature is explained by the presumed optimality of the answer G. If $\neg I$, the answer G would not be useful, hence the hearer can infer that I. This reasoning presupposes that the speaker knows that the pure propositional content of answer G will induce the hearer to go to the garage. In this paper, we are interested in examples where the analogue presupposition is not met, as in the following example:

(2) An email was sent to all employees that bus tickets for a joint excursion have been bought and are ready to be picked up. By mistake, no contact person was named. Hence, H asks one of the secretaries:
H: Where can I get the bus tickets for the excursion?
S: Ms. Müller is sitting in office 2.07. (M)
+> Bus tickets are available from Ms. Müller. (I)

In contrast to Example (1), it can not be assumed that the pure content of M will induce the inquirer to perform an optimal action. The difference between the answer G in (1) and M in (2) can be illustrated as follows: Assume in (1) that H finds a map with all petrol stations in town and notices that (G) *there is a garage round the corner*. This will be sufficient information to induce him to go to this garage. Now assume that, in (2), H finds a list with all office numbers of all employees, and reads there that (M) *Ms. Müller is sitting in office 2.07.* If there is no a priori link between M and Ms. Müller having bus tickets, i.e. if the two events are probabilistically independent, then what he reads will not induce H to go to office 2.07. For any reasonable definition of *relevance*, the answer M in (2) is *irrelevant* to the decision problem of H. It follows that the OA model, and the other mentioned models, cannot explain this example.

We will introduce a new pragmatic principle in order to explain the implicatures in examples like (2). We call it the *Principle of Optimal Completion.* The OA model tells us which answers a rational speaker can choose in accordance with his preferences and knowledge. Hence, if the speaker chooses a non–optimal answer, then either he is deviating from the pragmatic principles incorporated in the OA model or he is making a *mistake.* The core of our solution proceeds from the assumption that the hearer's interpretation strategy must be robust against small mistakes by the speaker. Being robust means that the hearer is able to repair these small mistakes and to *complete* under–informative sentences like M to sentences which would be optimal answers in the sense of the OA model.

In (2), we can assume that the optimal answer that S should have given is *Ms. Müller has the tickets. She is sitting in office 2.07* (F). The actual answer M is a sub-sentence of F. If the speaker follows the best strategy, then the OA model predicts that he can not answer M. Seen from within the model, using M is a *mistake*. Hence, in accordance with our core idea, we have to say what it means that a hearer strategy is robust against speaker's strategies which mix choosing M and F. If there is no other possibly optimal form F' such that M could be completed to F', then the hearer is safe to interprete M as a short form of F. Along these lines, we show that, from the assumption that the hearer's strategy is robust against small mistakes, it follows that there is only one way to interpret M, namely, as meaning F. This entails that the speaker can take advantage and produce, by intention or not, less costly answers, including apparently irrelevant answers. Thus the example can be explained.

In order to turn this sketch into a theory, we first of all have to spell out what we mean by *small* mistake and by a strategy being *robust* against them. As already mentioned, we model question–answering situations by the OA model [1], which concentrates on the pragmatically relevant parameters of the more general signalling games [5]. We derive a concept of robust interpretation strategies by (a strong) modification of the game theoretic notion of *trembling hand perfect* equilibria [10]. A strategy pair (s, h) is a trembling hand perfect equilibrium if each of the two strategies not only is a best response to the other one but also remains a best response if we add a small amount of noise to the other strategy. Our modification will refer i.a. to the kind of trembles that we allow.

The paper divides into two sections. In the first section, we introduce the OA model, which tells us how to calculate optimal answers and their implicatures. In the second section, we introduce the Principle of Optimal Completion. We will show that our model is able to handle scalar implicatures as well as the above mentioned relevance implicatures. In [3], the scalar implicature from *some* to *not all* can only be explained if *some but not all* has a higher a priori probability than *all*. The improved model will also predict the implicature in cases where *all* has the higher a priori probability.

2 The Optimal–Answer Model

It takes two for tango, and it takes two for a conversation. Conversation is characteristically a *cooperative effort* [4, p. 26]. Our contributions are not isolated sentences but normally subordinated to a joint purpose. In the Out-of-Petrol Example (1), the joint purpose is to solve the decision problem of where to go and look for petrol. 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* S, and the player who receives the answer the *inquirer* H. In game theory, the behaviour of agents is represented by *strategies*, i.e. functions that select actions for each of their possible knowledge states. The expert's action will always be an answer, the inquirer's action may e.g. be a decision about how to classify a a certain event, or, in the case of (1), where to look for petrol.

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 it is required that one assume a speaker to think in order to preserve the assumption that he is observing the Co-operative Principle (and perhaps some conversational maxims as well), ..." [4, p. 86]

In a game theoretic model, what the speaker utters is determined by his strategy s. If 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 [5, 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 [5], 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 $P : \Omega \to \mathbb{R}$ 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 $A \subseteq \Omega$ we set $P(A) = \sum_{v \in A} P(v)$. 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* $u : \mathcal{A} \times \Omega \to \mathbb{R}$ *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 to 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:

$$EU(a) = \sum_{v \in \Omega} P(v) \times u(a, v). \tag{2.1}$$

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 $P(H|A) := P(H \cap A)/P(A)$ for $P(A) \neq 0$. In terms of this conditional probability function, the *expected utility after learning* A is defined as $EU(a|A) = \sum_{v \in \Omega} P(v|A) \times u(a, v)$. H 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 H'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.

$$\mathcal{B}(A) := \{a \in \mathcal{A} \mid \forall b \in \mathcal{A}\ EU_H(b|A) \leq EU_H(a|A)\}. \tag{2.2}$$

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_S that represents his expectations about the world:

Definition 2 *A five–tuple $\sigma = \langle \Omega, P_S, P_H, \mathcal{A}, u\rangle$ is a* support problem *if (Ω, P_S) is a finite probability space and $D_\sigma = \langle(\Omega, P_H), \mathcal{A}, u\rangle$ a decision problem such that:*

$$\forall X \subseteq \Omega\ P_S(X) = P_H(X|K) \text{ for } K = \{v \in \Omega \mid P_S(v) > 0\}. \tag{2.3}$$

Condition (2.3) implies that the expert's beliefs cannot contradict the inquirer's expectations, i.e. for $A, B \subseteq \Omega$: $P_S(A) = 1 \Rightarrow P_H(A) > 0$.

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

Inquirer H Asks		Expert S answers		H decides for action		Evaluation
$\downarrow$		$\downarrow$		$\downarrow$		$\downarrow$
$\bullet$	$\xrightarrow{Q?}$	$\bullet$	$\xrightarrow{A}$	$\bullet$	$\xrightarrow{a}$	$\bullet$
$\uparrow$		$\uparrow$		$\uparrow$		$\uparrow$
expectations of H $\langle \Omega, P_H\rangle$		expectations of S $\langle \Omega, P_S\rangle$		expectations of H $\langle \Omega, P_H(\,.\,|A)\rangle$		utility measure $u(v, a)$

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

We add here a further Gricean maxim, the *Maxim of Quality*. We call an answer A *admissible* if $P_S(A) = 1$, i.e. if S believes A to be *true*. The Maxim of Quality is represented by the assumption that the expert S does only give admissible answers. For a support problem $\sigma = \langle \Omega, P_S, P_H, \mathcal{A}, u\rangle$ we set $Adm_\sigma := \{A \subseteq \Omega \mid P_S(A) = 1\}$. Hence, the set of optimal answers in σ is given by:

$$\mathrm{Op}_\sigma := \{A \in Adm_\sigma \mid \forall B \in Adm_\sigma\, EU_S(B) \leq EU_E(A)\}. \tag{2.4}$$

We write Op^h_σ if we want to make the dependency of Op on h explicit. In general, the solution to a support problem is not uniquely defined. Therefore, we introduce the notion of the *canonical* solution to a support problem.

Definition 3 *Let $\sigma = \langle \Omega, P_S, P_H, \mathcal{A}, u\rangle$ be a support problem. A* (mixed) strategy pair *for σ is a pair (s, h) such that s is a probability distribution over $\mathcal{P}(\Omega)$ and h a family of probability distributions $h(.|A)$ over $\mathcal{A}$. The* canonical solution *to σ is a pair (S, H) of mixed strategies which satisfy:*

$$S(A) = \begin{cases} |\mathrm{Op}_\sigma|^{-1},\ A \in \mathrm{Op}_\sigma \\ 0\ \textit{otherwise} \end{cases},\ H(a|A) = \begin{cases} |\mathcal{B}(A)|^{-1},\ a \in \mathcal{B}(A) \\ 0\ \textit{otherwise} \end{cases}. \tag{2.5}$$

We write $S(\,.\,|\sigma)$ if S is the function that maps each $\sigma \in \mathcal{S}$ to the speaker's canonical strategy, and $H(\,.\,|D_\sigma)$ if H is the function that maps the associated decision problem D_σ to the hearer's canonical strategy.

The expert may always answer everything he knows, i.e. he may answer $K := \{v \in \Omega \mid P_S(v) > 0\}$. From condition (2.3) it trivially follows that $\mathcal{B}(K) = \{a \in \mathcal{A} \mid \forall b \in \mathcal{A}\ EU_S(b) \leq EU_S(a)\}$. If expert and inquirer follow the canonical solution, then it is easy to see that:

$$\mathrm{Op}_\sigma = \{A \in Adm_\sigma \mid \mathcal{B}(A) \subseteq \mathcal{B}(K)\}; \tag{2.6}$$

We can call an answer A *misleading* if $\mathcal{B}(A) \not\subseteq \mathcal{B}(K)$; hence, (2.6) implies that Op_σ is the set of all non–misleading answers.

From now on, we will always assume that speaker and hearer follow the canonical solution.

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 I in a situation σ iff $P^\sigma_S(I) = 1$. 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. 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 I when making his utterance A, iff the speaker believes that I in all support problems which are indiscernible and in which A is an optimal answer. This leads to the following definition:

Definition 4 (Implicature) *Let $\mathcal{S}$ be a set of support problems and $\sigma \in \mathcal{S}$. Let $A, I \subseteq \Omega$ be two propositions with $A \in \mathrm{Op}_\sigma$. Then we set:*

$$A +> I \Leftrightarrow \forall \hat{\sigma} \in [\sigma]_{\mathcal{S}} \, (A \in \mathrm{Op}_{\hat{\sigma}} \rightarrow P_S^{\hat{\sigma}}(I) = 1), \tag{2.7}$$

with $[\sigma]_{\mathcal{S}} := \{\hat{\sigma} \in \mathcal{S} \mid D_\sigma = D_{\hat{\sigma}}\}$. If $A +> I$, then we say that the utterance of A implicates that I in σ.

As the hearer has to check all support problems in $[\sigma]_{\mathcal{S}}$, it follows that we find the more implicatures the smaller $\mathcal{S}$ is. We are especially interested in cases in which the speaker is a real expert. Let $O(a)$ be the set of all worlds in which a is an optimal action:

$$O(a) := \{w \in \Omega \mid \forall b \in \mathcal{A} \, u(w, a) \geq u(w, b)\}. \tag{2.8}$$

Then, we can say that the answering person is a real expert for a decision problem if he knows an action that is best in all possible worlds. We represent this information in $\mathcal{S}$ and arrive at the following criterion for implicatures:

Proposition 5 *Let $\mathcal{S}$ be a set of support problems such that $\forall \sigma \in \mathcal{S} \, \exists a \in \mathcal{A} \, P_S^\sigma(O(a)) = 1$. Let $\sigma \in \mathcal{S}$ and $A, I \subseteq \Omega$ be two propositions with $A \in \mathrm{Op}_\sigma$. Then, with $A^* := \{v \in \Omega \mid P_H(v) > 0\}$, it holds that:*

$$A +> I \text{ iff } \bigcap_{a \in \mathcal{B}(A)} O(a) \cap A^* \subseteq I. \tag{2.9}$$

For a proof see [2]. We use this criterion in the following examples.

2.3 Examples

We consider three examples: the Out–of–Petrol example, the Bus–Ticket example, and scalar implicatures. For more examples, we refer to [3]. We start with the Out–of-Petrol example (1). We distinguish three worlds $\{w_1, w_2, w_3\}$ and two actions $\{\text{go-to-g}, \text{search}\}$. G is the answer "*There is a garage round the corner*," and I the implicature "*The garage is open.*" The utilities and worlds are defined by the following table:

Ω	G	I	go-to-g	search
w_1	+	+	1	ε
w_2	+	−	0	ε
w_3	−	−	0	ε

The expert knows that he is in w_1. We assume that P_H and ε are such that $EU_H(\text{go-to-g}|G) > \varepsilon$, 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 $\mathcal{B}(G) = \{\text{go-to-g}\}$. We see that $O(\text{go-to-g}) = \{w_1\} = I$. Hence, by Lem. 5, it follows that $G \mathrel{+\!\!>} I$.

Now, we compare this situation with the slightly different Bus–Ticket example (2). The possible worlds in Ω differ according to whom the tickets can be picked up from, and according to the office number of this person. To simplify the model, we only consider four worlds and two actions. In the following table I stands for '*Bus tickets are available from Ms. Müller*', M for '*Ms. Müller is sitting in office 2.07*'. We assume that there are exactly two staff from whom bus tickets may be available, Ms. Müller and Mr. Schmidt, and that they are available from Ms. Müller iff they are not available from Mr. Schmidt. Furthermore, we assume that either staff is sitting in office 2.07 or 3.11, and that the one is sitting in office 2.07 iff the other one is sitting in 3.11. We assume that all possibilities are equally probable:

Ω	I	M	go-to-2.07	go-to-3.11
w_1	+	+	1	0
w_2	+	−	0	1
w_3	−	+	0	1
w_4	−	−	1	0

The expected utility of either action before learning anything is $\frac{1}{2}$, and after learning M the expected utilities still are $\frac{1}{2}$. Especially, if S knows that w_1, then M is not an optimal answer, and no implicatures are defined for it.

As a third example, we consider scalar implicatures. In (3), it has to be explained why $F_\exists$ implicates that not $F_\forall$:

(3) **a)** All of the boys came to the party. ($F_\forall$)
b) Some of the boys came to the party. ($F_\exists$)

We assume that Ω contains three worlds w_1, w_2, w_3. In w_1 all boys came, in w_2 some but not all, and in w_3 none came. The hearer's task is to find out what the actual world is. We only distinguish between success and failure. Hence we can identify the hearer's actions with the worlds w_i, and the expected utility of choosing w_i after learning proposition X with $P_H(w_i|X)$. If the hearer learns that $F_\exists$ and if $P_H(w_2) > P_H(w_1)$, then the set $\mathcal{B}(F_\exists)$ of optimal responses to $F_\exists$ is $\{w_2\}$. As $\{w_2\} \subseteq F_\exists \setminus F_\forall$, it follows with Lem. 5 that $F_\exists \mathrel{+\!\!>} \neg F_\forall$. But if $P_H(w_2) < P_H(w_1)$, then it would follow that $F_\exists \mathrel{+\!\!>} F_\forall$, which is contra-intuitive. We will see, that the Principle of Optimal Completion will make the implicature $F_\exists \mathrel{+\!\!>} \neg F_\forall$ independent of the hearer's expectations P_H.

3 The Principle of Optimal Completion

As mentioned in the introduction, we introduce a new pragmatic principle in order to explain examples like (2). This principle is motivated by the assumption that hearer's interpretation strategies must be robust against small mistakes by the speaker. In this context, we call any utterance a *mistake* if it is not predicted by the OA model. Obviously, it would not be reasonable to assume that the hearer can repair *any* mistake by the speaker. We only consider mistakes which consist in the production of *incomplete* utterances. An utterance is incomplete if the speaker had an optimal proposition in mind but only asserted a part of it. It is then left to the addressees to infer the full proposition, i.e. to complete the utterance to an optimal answer.

The general explanation of implicatures remains unchanged. We will especially not alter condition (2.7) in the previous definition of implicatures. The effect of optimal completion is a shift from the canonical hearer strategy H to a robust strategy $\bar{H}$, which in turn changes the set of optimal answers from which the speaker can make his choice. Hence, the shift from H to $\bar{H}$ will also lead to a shift from the canonical strategy S to a new speaker strategy $\bar{S}$. Implicatures are then calculated by using condition (2.7) relative to $(\bar{S}, \bar{H})$.

3.1 Optimal Completion and Efficient Clarification Requests

In the following, we need representations of answering situations which include explicit representations of linguistic forms and their meanings. We denote the set of forms by $\mathcal{F}$, and assume that there is a fixed semantic interpretation function $[\![\,.\,]\!]$ which maps forms F to propositions. Furthermore, we add a function $c : \mathcal{F} \to \mathrm{I\!R}^+ \setminus \{0\}$ that measures the costs of producing forms. We call a tuple $\langle \Omega, P_S, P_H, \mathcal{F}, \mathcal{A}, u, c, [\![\,.\,]\!] \rangle$ an *interpreted support problem with nominal costs* if $\langle \Omega, P_S, P_H, \mathcal{A}, u \rangle$ is a support problem which satisfies for all $F, H \in \mathcal{F}$: $EU_S([\![F]\!]) < EU_S([\![H]\!]) \Rightarrow EU_S([\![F]\!]) < EU_S([\![H]\!]) - c(H)$. That the costs of forms are *nominal* means that they are positive but very small, so small that they are always smaller than the positive differences of the expected utilities of the expressed propositions. This ensures that the answering expert will always choose an answer which expresses an optimal proposition.

Before introducing optimal completion, we make an addition to the basic optimal answer model which is crucial if optimal completion should not only explain implicatures of irrelevant answers but should also explain the scalar implicatures in (3). Let us consider an example similar to (2). Assume that a bike messenger H approaches the secretary S with a parcel and asks where to deliver it, and the secretary answers thereupon: '*It is for Ms. Müller.*' This information will not be sufficient if there are many offices and H doesn't know the building. The natural response of the messenger is a clarification request CR asking for the office of Ms. Müller. The request CR will lead to an answer which allows H to choose an optimal action a afterwards. In order to capture this possibility, we will add what we call a *efficient clarification request* CR to the hearer's action set. *Efficient* means here that its costs are nominal and its payoff high.[1] This has dramatic effects on the previous models due to backward

[1] This means that for all A $EU_H(\mathrm{CR}|A) = \sum_w P_H(w|A)u(a_w, w) - c(\mathrm{CR})$ with $a_w \in \{a \mid \forall b\, u(a, w) \geq u(b, w)\}$. Hence, if $\forall a \in \mathcal{A}\, P_H(O(a)|A) < 1$, then nominality of costs

induction. In a situation in which the speaker gives an answer $A = \{w_1, w_2\}$ which does not determine a unique optimal action, see Fig 1, the hearer has to make a risky choice. The existence of efficient clarification requests means that the hearer will always avoid this decision. Our previous models implicitly

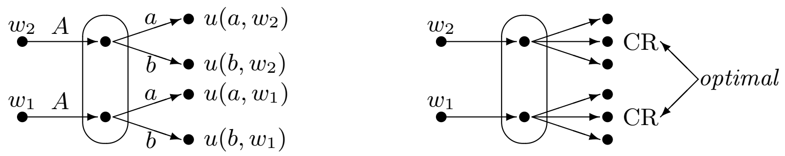

Fig. 1. Left: Without CR risky choice between a and b.

assumed that making clarification requests is not an option. If the hearer makes a clarification request, the answering expert has to produce an extra utterances. This leads to production costs which are higher than the cost of immediately producing an optimal answer. Hence, the speaker has an incentive to preempt the possibility of clarification requests. We show the effects on the basic OA model in our discussion of examples in Sec. 3.3.

Our definition of Implicatures in (2.7) implies that only optimal answers can have implicatures. As the definition of *optimal answer* depends on the hearer's strategy H, a change from the canonical strategy H to a robust strategy $\bar{H}$ will also change the set of utterances for which implicatures are predicted. If the speaker utters E, and E is a proper part of an optimal answer F, then the principle of optimal completion says that the hearer will complete E to F, i.e. interpret utterance E as an indicator of the speaker's intention to utter F. Let us write $E \lhd F$ for *utterance* E *is a proper part of* F. We assume that $\lhd$ is an undefined, primitive relation. There are obvious constraints that must be satisfied if the success of the principle of optimal completion is to be guaranteed. The triggering of the completion process must be unambiguous. This entails that the incomplete utterance must not be an optimal answer itself. For example, $E =$ '*all of the boys*' is a sub-form of $F =$ '*almost all of the boys*,' but an utterance of E should not trigger a completion to F. Furthermore, there must only be one optimal proposition to which the utterance can be completed. For example, in (2) there are many answers of which "*Ms. Müller is sitting in office 2.07*" is a sub–form. Not only "*Ms. Müller has the tickets. She is sitting in office 2.07*" but also e.g. of "*I don't know. Last time it was Ms. Müller who had the bus tickets. She is sitting in office 2.07.*" But these answers are optimal in different contexts. If it is common knowledge that the speaker knows the actual state of the world, then the last answer is ruled out as non-optimal.

entails that $\mathcal{B}(A) = \{\text{CR}\}$. If $\exists a \in \mathcal{A}\, P_H(O(a)|A) = 1$, then $\mathcal{B}(A) = \{a \,|\, P_H(O(a)|A) = 1\}$ and $O(\text{CR}) = \emptyset$ because $P_H(O(a)|A) = 1 \Rightarrow EU_H(\text{CR}|A) = EU_H(a|A) - c(\text{CR})$.

The concept that guides our game theoretic interpretation of the principle of optimal completion is the *trembling hand perfect* equilibrium, e.g. [7, Def. 248.1]. In the context of support problems, a trembling hand prefect equilibrium is a pair of mixed strategies (s, h) such that there exists a sequence $(s^k, h^k)_{k=0}^{\infty}$ of completely mixed strategies which converge to (s, h) such that s is a best responses to each h^k and h to each s^k. A strategy is *completely mixed* if it chooses every possible action with positive probability. That (s, h) is robust against *small* mistakes is captured by the condition that s and h need only to be best responses if h^k and s^k come close to h and s.

The criterion of trembling hand perfection asks of strategies to be robust against all kind of mistakes. We are interested in this robustness as an ability to repair mistakes which result from the production of sub–parts of optimal utterances. Hence, we will restrict *trembles* to sub-forms of optimal forms. We take into account the effect of clarification requests in the definition of *unique optimal completability* in Def. 7. There, we implicitly assume that the inquirer reacts to an ambiguous answer with a clarification request.

3.2 The Game Theoretic Model

We first define what it means that a form E is optimally completable to a form F. First, E must be a sub-form of F, both must be admissible, i.e. $P_s^\sigma([\![E]\!]) = P_s^\sigma([\![F]\!]) = 1$, and only F must be optimal. Furthermore, F has to be a minimal optimal form to which E can be completed. *Minimality* is here meant relative to a primitive sub–form relation $\lhd$. In the following definition, we denote by $\min_\lhd M$ the $\lhd$–minimal elements of M.

Definition 6 (Optimal Completion) *We say that, for a support problem σ, a form E can be* optimally completed *to form F, $oc(\sigma, E, F)$, iff $E \in Adm_\sigma \setminus \mathrm{Op}_\sigma$ and $F \in \min_\lhd \{F \in \mathrm{Op}_\sigma \mid E \lhd F\}$.*

This does not yet include the uniqueness condition. As the hearer does not know the support problem σ but only his decision problem D_σ, it must be excluded that E can be optimally completed to several different forms in support problems with the same associated decision problem D_σ.

Definition 7 (Unique Optimal Completion) *Let $\mathcal{S}$ be a given set of support problems with a joint decision problem. We say that E can be* uniquely optimally completed *to F, $uoc(E, F)$, if (1) $\exists \sigma \in \mathcal{S}\ oc(\sigma, E, F)$ and (2) for all $\sigma' \in \mathcal{S}$: $E \notin \mathrm{Op}_{\sigma'} \wedge \forall F' \in \mathcal{F}(oc(\sigma', E, F') \Rightarrow F = F')$.*

The uniqueness condition guarantees that the optimal super-form is recoverable from the non-optimal sub-form. As explained before, we only consider speaker's mistakes that are restricted to sub-forms which can be uniquely completed to optimal forms. Inspired by trembling hand perfection, we represent the possibility of speaker's mistakes by noisy strategies s^ϵ which approximate the canonical strategy s. If F is an optimal form for σ, and if the set of uniquely optimally completable sub–forms of F is not empty, then a speaker who follows

s^ϵ will choose one of these sub–forms with probability ϵ. If F doesn't have such sub–forms, then the probability of choosing F is the same for s^ϵ and s.

Definition 8 *Let $\mathcal{S}$ be a given set of support problems with a joint decision problem. Let $\sigma \in \mathcal{S}$, $F \in \mathrm{Op}_\sigma$, and n the cardinality of $\{E' \in \mathcal{F} \mid uoc(E', F)\}$. An* epsilon sub-form approximation *of a mixed speaker's strategy $s(\,.\,|\sigma)$ is a probability distribution $s^\epsilon(\,.\,|\sigma)$ on $\mathcal{F}$ such that (1) if $n = 0$, then $s^\epsilon(F|\sigma) = s(F|\sigma)$, and (2) if $n > 0$ and if E is such that $uoc(E, F)$, we set:*

1. *$s^\epsilon(F|\sigma) = (1 - \epsilon)s(F|\sigma)$,*
2. *$s^\epsilon(E|\sigma) = \epsilon\, n^{-1}\, s(F|\sigma)$,*

For all other forms E, $s^\epsilon(E|\sigma) = 0$.

Hence, $s^\epsilon(E|\sigma) > 0$ iff $E \in \mathrm{Op}_\sigma \vee \exists F \in \mathrm{Op}_\sigma\; uoc(E, F)$. Due to the uniqueness condition, the hearer's best response $\bar{H}$ to these noisy speaker strategies can easily be found. We call it the *sub-form extension* of the canonical solution H:

Definition 9 *Let $\mathcal{S}$ be a given set of support problems with a joint decision problem $\langle(\Omega, P_{H}), \mathcal{A}, u\rangle$. Let (S, H) be the canonical solution to $\sigma \in \mathcal{S}$. Then, the* sub-form extension *$\bar{H}$ of H is defined as follows:*

1. *If $F \in \bigcup_{\sigma\in\mathcal{S}} \mathrm{Op}_\sigma$, then $\bar{H}(a|F) = H(a|F)$.*
2. *If E, F are such that $uoc(E, F)$, then $\bar{H}(a|E) = H(a|F)$.*

To all forms E for which there is no $F \in \bigcup_\sigma \mathrm{Op}_\sigma$ such that $E = F$ or $uoc(E, F)$, we assume that the hearer reacts with a clarification request.

The following lemma shows that the sub-form extension $\bar{H}$ provides a choice of action for all answers which the speaker may choose with positive probability, and that all the choices are optimal. This holds for the ϵ sub-form approximations s^ϵ, as well as for the canonical strategy s itself.

Proposition 10 *Let $\mathcal{S}$ be a given set of support problems with a joint decision problem $\langle(\Omega, P_{H}), \mathcal{A}, u\rangle$. Let (S, H) be the canonical solution to $\mathcal{S}$ and $\bar{H}$ the sub-form extension of H. For $\sigma \in \mathcal{S}$ let $K_\sigma := \{v \in \Omega \mid P_{S}^{\sigma}(v) > 0\}$. Then, it holds for all ϵ and all forms E with $s^\epsilon(E|\sigma) > 0$ that (1) $\exists a \in \mathcal{A}\; \bar{H}(a|E) > 0$, and (2) $\forall a \in \mathcal{A}\; (\bar{H}(a|E) > 0 \Rightarrow a \in \mathcal{B}(K_\sigma))$.*

Proof: The first proposition holds by definition of $\bar{H}$. Let $\sigma \in \mathcal{S}$. Let $H(a|E) > 0$ and $s^\epsilon(E|\sigma) > 0$. Then, $E \in \mathrm{Op}_\sigma^H$ or there exists $F \in \mathrm{Op}_\sigma^H$ such that $uoc(E, F)$. If $E \in \mathrm{Op}_\sigma^H$, then $\bar{H}(a|E) = H(a|E)$, hence $a \in \mathcal{B}(K_\sigma)$ by (2.5) and (2.6). If there exists $F \in \mathrm{Op}_\sigma$ such that $uoc(E, F)$, then $\bar{H}(a|E) = H(a|F)$, therefore again $\bar{H}(a|E) > 0 \Rightarrow a \in \mathcal{B}(K_\sigma)$. ■

As $s^\epsilon(E|\sigma) > 0$ iff $E \in \mathcal{F}_\sigma = \{E \mid \exists F \in \mathrm{Op}_\sigma(E = F \vee uoc(E, F))\}$, it follows that $s^\epsilon(E|\sigma) > 0$ implies that E is a non-misleading answer, see (2.6), hence, the speaker can optimise his strategy by choosing answers from $\mathcal{F}_\sigma$ which have minimal costs. If we assume that the speaker prefers forms which are minimal relative to the sub-form relation $\lhd$, then the set of speaker–optimal answers relative to $\bar{H}$ is the set of $\lhd$-minimal elements of $\mathcal{F}_\sigma$, which we denote by $\mathrm{Op}_\sigma^{\bar{H}}$. Let $\bar{S}$ be the speaker strategy which chooses the elements of $\mathrm{Op}_\sigma^{\bar{H}}$ with equal probability. We call it the *sub-form extension* of the canonical strategy S. Disregarding nominal costs of forms, it is clear by construction that $(\bar{S}, \bar{H})$ and all $(s^\epsilon, \bar{H})$ are (weakly) dominating all other solutions.

With these preparations, we now can represent the *Principle of Optimal Completion*. It just means that speaker and hearer follow the sub-form extension $(\bar{S}, \bar{H})$ of the canonical solution (S, H). The definition of implicatures remains unchanged. If $\mathcal{S}$ is a set of interpreted support problems with a common decision problem, then, by adjusting (2.7) to $(\bar{S}, \bar{H})$, we arrive at:

$$A \mathrel{+\!\!>} I \Leftrightarrow \forall \sigma \in \mathcal{S}\, (A \in \mathrm{Op}_\sigma^{\bar{H}} \Rightarrow P_S^\sigma(I) = 1). \tag{3.10}$$

3.3 Examples

We again consider the examples from Sec. 2.3 and provide models that explain their implicatures using the principle of optimal completion. We start with the standard scalar implicatures:

(4) **a)** All of the boys came to the party. ($F_\forall$)
b) Some of the boys came to the party. ($F_\exists$)
$+\!\!>$ Some but not all of the boys came to the party. ($F_{\exists\neg\forall}$)

As in our discussion of Example 3, we assume that Ω contains two worlds w_1, w_2. In w_1, all boys came, and in w_2 some but not all. Here, and in the following examples, we assume that $P_H(w_i) > 0$ for $i = 1, 2$. The hearer's task is to find out what the actual world is. We again only distinguish between success and failure and identify the hearer's actions with the worlds w_i. Hence, the expected utility of choosing w_i after learning proposition X is $P_H(w_i|X)$. Let the hearer's decision problem $\langle(\Omega, P_H), \mathcal{A}, u\rangle$ be any decision problem that satisfies these conditions.

We saw in Sec. 2.3 that $P_H(w_2) \leq P_H(w_1)$ entails that $F_\exists$ does not implicate $F_{\exists\neg\forall}$. We show now that the principle of optimal completion implies that the implicature becomes independent of $P_H(w_1)$. For this, we have to assume common knowledge of the fact that the answering expert S knows the actual world w_i. We encode common knowledge in the background set $\mathcal{S}$ of possible support problems. Therefore, we assume that $\mathcal{S}$ contains exactly two support problems σ_1 and σ_2 with $P_S^{\sigma_i}(w_i) = 1$. For defining the full interpreted support problems $\langle\Omega, P_S, P_H, \mathcal{F}, \mathcal{A}, u, c, [\![\,.\,]\!]\rangle$, it remains to define the speaker's set of forms $\mathcal{F}$, their meanings and the cost function. Let $\mathcal{F} = \{F_\forall, F_{\exists\neg\forall}, F_\exists\}$ and $[\![F_\forall]\!] = \{w_1\}$, $[\![F_{\exists\neg\forall}]\!] = \{w_2\}$, and $[\![F_\exists]\!] = \{w_1, w_2\}$. We assume that the costs reflect the sub-form relation $\lhd = \{\langle F_\exists, F_{\exists\neg\forall}\rangle\}$. The following tables show the optimal answers for i) $P_H(w_1) < P_H(w_2)$, ii) $P_H(w_1) = P_H(w_2)$, and iii) $P_H(w_1) > P_H(w_2)$. Op_{σ_i} is the set of optimal answers which we derive from the basic OA model in Sec. 2.1; $\mathrm{Op}_{\sigma_i}^{+}$ is the set of optimal answer which we get if we add efficient clarification requests to the model; and $\mathrm{Op}_{\sigma_i}^{++}$ shows the effect of optimal completion. As $\mathrm{Op}_{\sigma_i}^{+}$ and $\mathrm{Op}_{\sigma_i}^{++}$ are identical for all three cases, we depict them only once.

i) σ_i	Op_σ	$\mathrm{Op}^+_{\sigma_i}$	$\mathrm{Op}^{++}_{\sigma_i}$
w_1	$\{F_\forall\}$	$\{F_\forall\}$	$\{F_\forall\}$
w_2	$\{F_\exists, F_{\exists\neg\forall}\}$	$\{F_{\exists\neg\forall}\}$	$\{F_\exists\}$

ii) σ_i	Op_σ
w_1	$\{F_\forall\}$
w_2	$\{F_{\exists\neg\forall}\}$

iii) σ_i	Op_σ
w_1	$\{F_\forall, F_\exists\}$
w_2	$\{F_{\exists\neg\forall}\}$

In case i), $F_{\exists\neg\forall}$ is in Op_σ only if we ignore the speaker's preferences for short forms. If we include them, then only $F_\exists$ is optimal. In both cases, however, $F_\exists \in \mathrm{Op}_\sigma$, as we have seen in Sec. 2.3. But this holds only if efficient clarification requests are not available as by assumption $EU_H(w_i|[\![F_\exists]\!]) = P_H(w_i) < 1 - c(\mathrm{CR}) = EU_H(\mathrm{CR}|[\![F_\exists]\!])$. Hence, only $F_{\exists\neg\forall}$ is optimal once we take efficient clarification requests into account. Their availability results in the same optimal answers Op^+_σ in all three cases i)–iii). Clearly, in all cases, $F_\exists \in Adm_{\sigma_2} \setminus \mathrm{Op}^+_{\sigma_2}$, $F_\exists \lhd F_{\exists\neg\forall}$, and for all $i \neq 2$: $F_\exists \notin \mathrm{Op}^+_{\sigma_i}$ and $\neg\exists F' \in \mathrm{Op}^+_{\sigma_i}\ oc(\sigma_i, F_\exists, F')$. Hence, the uniqueness conditions, Def. 7, are satisfied, therefore $uoc(F_\exists, F_{\exists\neg\forall})$. By definition, it follows that $\bar{H}(w_i|F_\exists) = \bar{H}(w_i|F_{\exists\neg\forall})$. Hence, the addressee will choose w_2 after receiving $F_\exists$ which shows that $F_\exists \in \mathrm{Op}^{++}_{\sigma_i}$. By definition of $\bar{S}$, $\bar{S}(F_\exists|\sigma_2) = S(F_{\exists\neg\forall}|\sigma_2) = 1$. As $\bar{S}(F_\exists|\sigma_1) = S(F_\exists|\sigma_1) = 0$, it follows that $\bar{S}(F_\exists|\sigma_i) > 0 \Rightarrow P_S^{\sigma_i}(w_2) = 1$. Hence, $F_\exists$ implicates that not all boys came to the party.

We now turn to the Bus Ticket example (2). We consider the same model as in Sec. 2.3, p. 8, where we assumed that there are exactly two staff from whom bus tickets may be available, Ms. Müller and Mr. Schmidt, that they are available from Ms. Müller iff they are not available from Mr. Schmidt, and that one of them is sitting in office 2.07 iff the other one is sitting in 3.11. With the sentence frames $A(i,n)$ = 'i is sitting in office n,' and $B(i)$ = 'Bus tickets are available from i', we can describe the speaker's set of forms $\mathcal{F}$ from which he can choose as the set of sentences of the form $B(i)$, $A(i,n)$, or $B(i) \wedge A(i,n)$, with their meaning defined in the usual way. With $i = 0$ for Ms. Müller, and $i = 1$ for Mr. Schmidt, the possible worlds and payoffs can be read off from the first columns of the following table.

Ω	$B(0)$	$A(0, 2.07)$	go-to-2.07	go-to-3.11	Op_{w_j} $(= \mathrm{Op}^+_{w_j})$
w_1	$+$	$+$	1	0	$B(0) \wedge A(0, 2.07)$
w_2	$+$	$-$	0	1	$B(0) \wedge A(0, 3.11)$
w_3	$-$	$+$	0	1	$B(1) \wedge A(1, 3.11)$
w_4	$-$	$-$	1	0	$B(1) \wedge A(1, 2.07)$

The sub-form relation $\lhd$ is defined in the obvious way. Again, we have to assume that the answering expert knows the actual state of the world. In the scenario of (2), w_1 is the actual world. The optimal answers can be seen in the last column of the table.

$M := A(0, 2.07)$ is a sub-form of the optimal answer $B(0) \wedge A(0, 2.07) =: I \wedge M$. In w_1, S believes both to be true, and there is no other world w_j where these conditions are satisfied for M. This means that $uoc(M, I \wedge M)$. It follows with Def. 9 that $\bar{H}(\text{go-to-2.07}|M) = 1$ and that $M \in \mathrm{Op}^{\bar{H}}_{\sigma_i}$ implies $i = 1$ and $P_S^{\sigma_i}(I) = 1$. Hence, by (3.10), $M +\!> I$. This proves the claim.

Finally, we turn to the *Out-of-petrol* example (1) and reconsider the model of Sec. 2.3. If we add efficient clarification requests to the model, then answer

G is not optimal any more. To see this, we have to add some more detail to the model. We assume that there are two garages g_1 and g_2, and define Ω, actions, and propositions as in the table below. Let's assume that G_1 corresponds to the assertion "*There is a garage round the corner to the left,*" and G_2 to "*There is a garage round the corner to the right.*" In (1), it is implicitly assumed that it is common knowledge that the speaker knows the actual world. With this assumption, it follows from $EU_H(\text{go-to-g}_i|G_i) = P_H(G_i \wedge I_i) < 1 - c(\text{CR}) = EU_H(\text{CR}|G_i)$ that $G_i \notin \text{Op}^+_{w_j}$. In the table below, $G_i \wedge I_i$ is an element of Op_{w_j} and $\text{Op}^{++}_{w_j}$ only if we do not take into account the speaker's preferences regarding forms. G_1 can be optimally completed to $G_1 \wedge I_1$ in w_1, w_2, and w_3 but not in w_4 and w_5, see $\text{Op}^+_{w_j}$. By definition, $\bar{S}(G_1|w_j) > 0$ iff $j \in \{1,2,3\}$, and therefore $G_1 +> I_1$.

Ω	G_1	I_1	G_2	I_2	g-t-g$_1$	g-t-g$_2$	srch	Op_{w_j}	$\text{Op}^+_{w_j}$	$\text{Op}^{++}_{w_j}$
w_1	+	+	+	+	1	1	ε	$\{G_i, G_i \wedge I_i\}$	$\{G_i \wedge I_i\}$	$\{G_i, G_i \wedge I_i\}$
w_2	+	+	+	−	1	0	ε	$\{G_1, G_1 \wedge I_1\}$	$\{G_1 \wedge I_1\}$	$\{G_1, G_1 \wedge I_1\}$
w_3	+	+	−	−	1	0	ε	$\{G_1, G_1 \wedge I_1\}$	$\{G_1 \wedge I_1\}$	$\{G_1, G_1 \wedge I_1\}$
w_4	+	−	+	+	0	1	ε	$\{G_2, G_2 \wedge I_2\}$	$\{G_2 \wedge I_2\}$	$\{G_2, G_2 \wedge I_2\}$
w_5	−	−	+	+	0	1	ε	$\{G_2, G_2 \wedge I_2\}$	$\{G_2 \wedge I_2\}$	$\{G_2, G_2 \wedge I_2\}$

References

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

[2] A. Benz (2008): *How to Set Up Normal Optimal Answer Modells;* ms., ZAS Berlin.

[3] 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 .

[4] H.P. Grice (1989): *Studies in the Way of Words;* Harvard University Press, Cambridge Massechusetts.

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

[6] A. Merin (1999): *Information, Relevance, and Social Decisionmaking: Some Principles and Results of Decision-Theoretic Semantics;* in: L.S. Moss, J. Ginzburg, and M. de Rijke (eds.): *Logic, Language, and Information* Vol. 2; Stanford.

[7] M.J. Osborne, A. Rubinstein (1994): *A Course in Game Theory;* The MIT Press.

[8] P. Parikh (2001): *The Use of Language;* CSLI Lecture Notes, Stanford, CA.

[9] R. v. Rooij (2004): *Utility of Mention-Some Questions;* in *Research on Language and Computation* 4(3), pp. 401-416.

[10] R. Selten (1975): *Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games;* in *International Journal of Game Theory* 4, pp. 25–55.