Outline of the Foundations for a Theory of Implicatures

Anton Benz
Centre for General Linguistics, Berlin

In this paper, we outline the foundations of a theory of implicatures. It divides into two parts. The first part contains the base model. It introduces signalling games, optimal answer models, and a general definition of implicatures in terms of natural information. The second part contains a refinement in which we consider noisy communication with efficient clarification requests. Throughout, we assume a fully cooperative speaker who knows the information state of the hearer. The purpose of this paper is *not* the study of examples. Our concern is the framework for doing these studies.

1 Introduction

Communication poses a coordination problem. We represent this coordination problem by signalling games (Lewis, 2002). The solutions to the coordination problem are strategy pairs which describe the speaker's signalling and the hearer's interpretation behaviour. The behaviour is an objective natural regularity, and the speaker's and hearer's strategies determine with which probability they will choose their respective actions given their respective information states. As natural regularity, the communicative process can be described as a causal Bayesian network (Pearle, 2000). From this representation, we derive the notion of *natural information* which is related to Grice' (1957) concept of *natural meaning*. We claim that this is a key concept for the understanding of pragmatics.

Natural information is *objective* information, i.e. it exists independently of the beliefs and intentions of language users. To justify this interpretation we have to interpret the probabilities in signalling games as *objective relative frequencies*. From this objective level we distinguish a subjective cognitive level at which probabilities are interpreted as *subjective probabilities*. We describe the

subjective level by *optimal answer* (OA) models. We justify this representation by a discussion of the theory of mind as incorporated in *iterated best response models* (Franke, 2009).

Accordingly, the first part of the paper divides into five sections. The first section introduces signalling games, the second section the concept of natural information and the general definition of implicature, and the third section the optimal answer models and their *canonical* solutions. The third section also discusses the relation between OA and iterated best response models. The fourth section applies the general definition of implicatures to OA models and signalling games. In Section 2, we present a lemma which provides us with a criterion for deciding whether or not a strategy pair is an objective Pareto Nash equilibrium of a signalling game. This lemma, Lemma 2.3 will play an important role in our discussion of aspects of bounded rationality, the theory of mind, and the *objective* justification of canonical solutions to OA models. The last section of the first part provides the proof of this lemma.

The second part of this paper starts out with a discussion of the idea that ambiguities are resolved by choosing the more probable interpretation, and that, as a consequence, the more probable interpretation of an ambiguous utterance is communicated with *certainty*. This principle figures prominently in Prashant Parikh's (2001) approach to game theoretic pragmatics, which basically assumes that all pragmatic strengthening and weakening of interpretation can be reduced to cases of disambiguation. We argue that the natural hearer's reaction to an ambiguity is to ask a clarification request. Hence in Section 8, we consider signalling games for which the hearer's action set contains efficient clarification requests. *Efficiency* means that clarification requests have nominal costs and lead to almost maximal payoffs. The availability of efficient clarification requests changes the equilibria of signalling games if we allow for *noisy* speaker strategies. This noise may have *external* causes, i.e. the kind of noise might not be predictable from game theoretic parameters. Hence, we introduce a very general model for representing noisy speaker strategies. This is done in Section 9. In this section, we also show how the canonical solutions to OA models change, and how the notion of implicatures applies to models representing noisy speaker strategies. Section 10, contains further characterisations of the equilibrium properties of canonical solutions for noisy games and the proof of a lemma analogous to Lemma 2.3. The final section contains some clarifications concerning our concept of *nominal* costs.

2 Signalling Games

Grice (1989, p. 26) characterised conversation as a *cooperative effort*. This means that the contributions of the interlocutors are not isolated sentences but subordinated to a joint purpose. In this paper, we will always assume that each assertion answers an implicit or explicit question by the hearer which in turn is embedded in a decision problem. The decision problem is such that the hearer has to make a choice between several actions. The hearer's choice of actions depends on his preferences regarding the actions' outcomes and his knowledge about the world. The speaker's message helps the inquirer in making his choice. The quality of a message depends on the action to which it will lead. Hence, communication poses a coordination problem to speaker and hearer. The speaker has to choose his contribution such that it induces the hearer to choose an optimal action; and the hearer has to consider the speaker's message and use the communicated information for making the best choice. We represent these coordination problems as *signalling games* (Lewis, 2002). The signalling games are such that first nature chooses a world v with probability $P(v)$; then again nature chooses a type θ, i.e. an information state, for the speaker S with conditional probability $p(\theta|v)$; then the speaker chooses a signal F with conditional probability $S(F|\theta)$, and finally the hearer chooses an act a with conditional probability $H(a|F)$. A branch of this game is depicted in the following figure:

$$\bullet \xrightarrow{P(v)} \underset{v}{\bullet} \xrightarrow{p(\theta|v)} \underset{S}{\bullet} \xrightarrow{S(F|\theta)} \underset{H}{\bullet} \xrightarrow{H(a|F)} \bullet \; u(v,\theta,F,a)$$

We formally define the signalling games as follows:

Definition 2.1 (Signalling Game) *A tuple* $\langle \Omega, \Theta, P, p, \mathscr{F}, \mathscr{A}, u \rangle$ *is a* signalling game *if:*

1. Ω *and* Θ *are non–empty finite sets;*
2. $P(\,.\,)$ *is a probability distribution over* Ω*;*
3. $p(\,.\,|v)$ *is a probability distribution over* Θ *for every* $v \in \Omega$*;*
4. $\mathscr{F}$ *and* $\mathscr{A}$ *are respectively the speaker's and hearer's action sets;*
5. $u : \Omega \times \Theta \times \mathscr{F} \times \mathscr{A} \to \mathbb{R}$ *is a shared utility function.*

We assume that $u(v,\theta,F,a)$ *can be decomposed into a difference* $u(v,a) - c(F)$ *for some real valued function* $u(v,a)$ *and a positive value* $c(F)$*.*

We assume that the general game structure is common knowledge. The speaker, in addition, knows θ when choosing signal F, and the hearer knows F when choosing action a. This means that the agents' strategies are functions of the following form:

- For each type $\theta \in \Theta$, the speaker's strategy $S(.|\theta)$ is a probability distribution over $\mathscr{F}$;
- For each signal $F \in \mathscr{F}$, the hearer's strategy $H(.|F)$ is a probability distribution over $\mathscr{A}$.

In principle, the probabilities could be interpreted as objective frequencies or as subjective probabilities. For reasons which will become clear in the next section, we interpret all the probabilities related to signalling games as objective frequencies.

Next, we introduce the notion of a *Nash equilibrium*. The speaker's expected utility $\mathscr{E}(S|H)$ of strategy S given a hearer strategy H is defined as:

$$\mathscr{E}(S|H) = \sum_{v\in\Omega} P(v) \sum_{\theta\in\Theta} p(\theta|v) \sum_{A\in\mathscr{F}} S(F|\theta) \sum_{a\in\mathscr{A}} H(a|F)\, u(v,\theta,F,a). \quad (2.1)$$

As the basic signalling games defined in Def. 2.1 are games of pure coordination, i.e. games in which the utility functions of both agents are identical, it follows that $\mathscr{E}(S|H) = \mathscr{E}(H|S)$. With these notions at hand, we can define:

Definition 2.2 (Nash Equilibrium) *A strategy pair (S,H) is a* Nash equilibrium *of a signalling game $\langle \Omega, \Theta, P, p, \mathscr{F}, \mathscr{A}, u\rangle$ iff:*

1. *For all speaker strategies S': $\mathscr{E}(S'|H) \leq \mathscr{E}(S|H)$,*
2. *For all hearer strategies H': $\mathscr{E}(H'|S) \leq \mathscr{E}(H|S)$,*

The equilibrium is strict *if we can replace $\leq$ by $<$. It is* weak *if it is not strict.*

For a game of pure coordination, a Nash equilibrium is a *Pareto Nash equilibrium* iff for all other Nash equilibria (S',H'): $\mathscr{E}(S'|H') \leq \mathscr{E}(S|H)$. In this case, we also say that (S,H) *(weakly) Pareto dominates* (S',H').

The textbook equilibrium concept for signalling games is the concept of a *Bayesian perfect equilibrium*. Bayesian perfection takes the player's information set into account. The player's strategy must be optimal given the information available to him at the time when he actually makes the decision. For the hearer, this is after receiving an answer F. Apart from the possible semantic meaning of the answer, the hearer is gaining additional information from the fact that the answer was given. Hence, the probability distribution that enters in the hearer's decision making is his prior distribution updated with the information gained by learning that a certain answer has been given. But, for the basic

signalling games which we consider, Baysian perfect equilibria and Nash equilibria in the sense of Definition 2.2 coincide. Although their definition is more complicated, it can be easier to do calculations for Bayesian perfect equilibria. We will do this in Section 6.

In general, it is often convenient or necessary to formulate constraints and do calculations with conditional probabilities, and not with P and p directly. The probability with which nature assigns type θ to speaker S in world v equals $P(v)\,p(\theta|v)$. Hence, the speaker's probability $\mu_S(v|\theta)$ for a world v after receiving type θ is a conditional probability defined as the probability to receive θ in v divided by the overall probability of receiving θ; see (2.2). For the hearer, we find an analogous probability distribution. He acts after receiving a signal F. Hence, the hearer's probability $\mu_H(v|F)$ of a world v after receiving F is the probability of receiving F in v divided by the overall probability of receiving signal F (2.2). The explicit definitions are as follows:

$$\mu_S(v|\theta) = \frac{P(v)\,p(\theta|v)}{\sum_w P(w)\,p(\theta|w)}, \quad \mu_H(v|F) = \frac{P(v)\,\sum_\theta p(\theta|v)\,S(F|\theta)}{\sum_w P(w)\,\sum_\theta p(\theta|w)\,S(F|\theta)}. \tag{2.2}$$

Here and in the following, we assume that the denominators are non–zero. For μ_S this means that there exists a w such that $P(w)\,p(\theta|w) > 0$, and for μ_H that there are w and θ for which $P(w)\,p(\theta|w)\,S(F|\theta) > 0$.

In later sections, we will often make use of the following abbreviations:

$$\mu_\Theta(\theta) := \sum_w P(w)\,p(\theta|w), \text{ and } \mu_{\mathscr{F}}(F) := \sum_w P(w) \sum_\theta p(\theta|w)\,S(F|\theta). \tag{2.3}$$

$\mu_{\mathscr{F}}(F)$ is the probability for the speaker producing F, and $\mu_\Theta(\theta)$ is the probability for the speaker's type to be θ. As it is clear from the argument which measure is meant, we will write $\mu(F)$ instead of $\mu_{\mathscr{F}}(F)$, and $\mu(\theta)$ instead of $\mu_\Theta(\theta)$.

Given type θ, the (speaker's) *expected utility* of an action a is defined by:

$$\mathscr{E}_S(a|\theta) = \sum_v \mu_S(v|\theta)\,u(v,a) \tag{2.4}$$

Similarly, given answer F, the (hearer's) *expected utility* of an action a is defined by:

$$\mathscr{E}_H(a|F) := \sum_v \mu_H(v|F)\,u(v,\theta,F,a). \tag{2.5}$$

The speaker's expected utility of a strategy S given his type θ is then:

$$\mathscr{E}_S(S|\theta) = \sum_A S(F|\theta) \sum_a H(a|F)\,\mathscr{E}_S(a|\theta) \tag{2.6}$$

And the hearer's expected utility of a strategy H given his information state after receiving signal F is then:

$$\mathscr{E}_H(H|F) := \sum_a H(a|F)\,\mathscr{E}_H(a|F) \tag{2.7}$$

We are now interested in a simple criterion for deciding whether a strategy pair is a Pareto Nash equilibrium. The criterion will only depend on S, H and the following set $\mathscr{B}(\theta)$ which is the set of all actions with maximal expected utility:

$$\mathscr{B}(\theta) = \{a \in \mathscr{A} \mid \forall b \in \mathscr{A}\ \mathscr{E}_S(b|\theta) \leq \mathscr{E}_S(a|\theta)\}. \tag{2.8}$$

Throughout the paper, we will make extensive use of the following fundamental lemma:

Lemma 2.3 *Let $\langle \Omega, \Theta, P, p, \mathscr{F}, \mathscr{A}, u\rangle$ be a signalling game. Let Θ^* be the set of all types θ for which $\exists v\, P(v)\, p(\theta|v) > 0$. Let (S,H) be a strategy pair which satisfies the following condition:*

$$\forall F \in \mathscr{F} \forall \theta \in \Theta^*\, (S(F|\theta) > 0 \Rightarrow H(\mathscr{B}(\theta)|F) = 1). \tag{2.9}$$

Then (S,H) is a Pareto Nash equilibrium. Furthermore, if H' is such that

$$\exists F \in \mathscr{F} \exists \theta \in \Theta^*\, \exists a \notin \mathscr{B}(\theta)\, (S(F|\theta) > 0 \wedge H'(a|F) > 0), \tag{2.10}$$

Then (S,H') is not a Nash equilibrium, in particular, it is $\mathscr{E}(H'|S) < \mathscr{E}(H|S)$.

We will prove this lemma in Section 6

3 Natural Information

In (1957), Grice introduced the distinction between *natural meaning* and *communicated meaning*. Natural meaning is the information which can be carried by an event or object independently of the beliefs and intentions of any person who may use this event or object for the purposes of communication. Grice used the following example for illustrating the concept of *natural meaning*:

(1) **a)** Those spots mean measles.

b) Those spots didn't mean anything to me, but to the doctor they meant measles.

In both sentences, the word *meaning* refers to natural meaning. The spots carry the information that the patient is infected with measles independently of any person using the spots for communicating that he is infected with measles, e.g. by pointing at the patient and saying: '*Look what he has!*.' The spots carry their

information due to a causal relation that exists between the infection and red spots on the skin. This causal relation is a natural regularity which is the basis for the inference from *red spots* to *measles*.

Causal relations can be represented by *causal networks*. The diagram in Figure 1 from (Pearle, 2000, p. 15) may serve as an illustration. $\mathscr{X}_0, \ldots, \mathscr{X}_4$

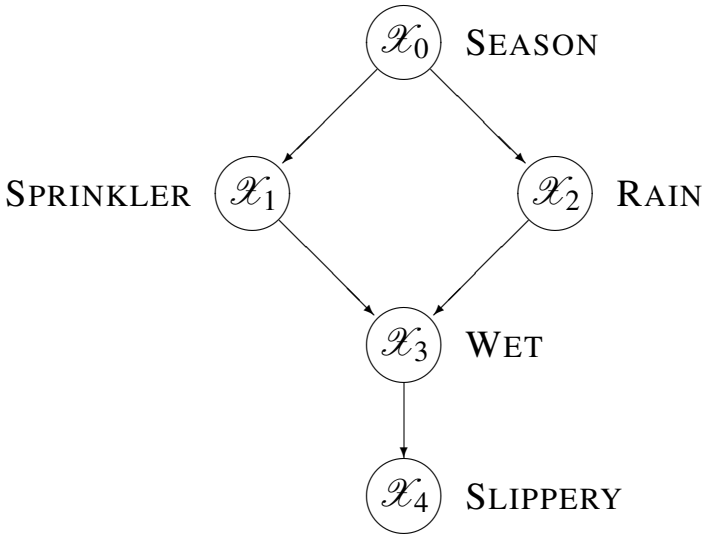
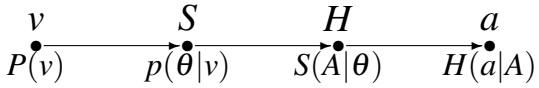


Fig. 1: A causal network.

are random variables which represent the state of the season and of a sprinkler, whether it rains, and whether a certain place is wet or slippery. The random variable for the season can take four different values, whereas the random variables for the sprinkler, the rain, and the wetness and slipperiness are only taking the Boolean values *true*, or *false*. In causal Bayesian networks, the causal dependencies are represented by *conditional probabilities* which hold between random variables. Given, e.g., that the slipperiness of a road is determined by its wetness, which in turn is determined by the fact whether a sprinkler is on, or whether it is raining, and that for example the state of the sprinkler is determined by the season, then we could say that: '*That the street is slippery means that the sprinkler was on or that it rained;*' or '*That the sprinkler is on means that it is summer*'. In both cases, the word *means* refers to natural meaning.

We now turn to the communication process. As we have seen in the last section, the context of communication can be described by the state of the world v, the speaker's information state θ, and a fixed information state of the hearer. Let Ω be the set of all possible worlds, and Θ of all possible speaker states. Again as in the last section, we identify the communicative behaviour of speaker and hearer with strategies S and H, i.e. with functions S which map the speaker's possible information states θ to probability distributions over a set $\mathscr{F}$ of possible utterances, and functions H which map utterance F to probability distributions over a set of hearer actions $\mathscr{A}$. Hence, S only depends on the speaker's

information state θ, and the hearer's strategy on the signal F which he receives from the speaker. We write $P(v)$ for the probability of a world v, and $p(\theta|v)$ for the probability of the speaker's information state θ given v. If P, p, S, and H are given, then we can think of the communicative process as a *Markovian* process, i.e. a process in which the probability of each successor state only depends on the predecessor states. A branch in this process is shown in the following graph:



In generally, we can think of the Ω, Θ, $\mathscr{F}$, and $\mathscr{A}$ as random variables in a causal Bayesian network in which the conditional probabilities P, p, S, and H define causal dependencies between these variables. Clearly, this identification assumes that all probabilities are objective frequencies. This is all we need to introduce a meaningful definition of *natural information*.

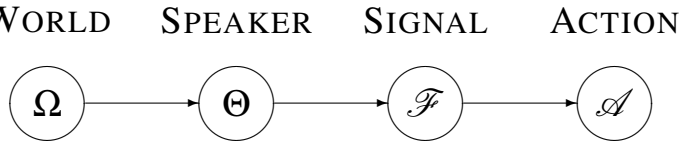


Fig. 2: The causal network associated to a signalling game.

For the following definitions, we abstract away from all particularities of linguistic communication. In order to make our definition not too far removed from our applications, we consider only graphs which represent a linear sequence of causal dependencies. But our definitions will immediately generalise to any causal Bayesian network which is represented by a directed acyclic graph. A linear graph of length $n+1$ is given by a pair $(\mathscr{X}_i, p_i)_{i=0,\ldots,n}$ for which:

1. $(\mathscr{X}_i)_{i=0\ldots,n}$ is a family of non-empty sets,
2. $p_0(\,.\,)$ a probability distribution over $\mathscr{X}_0$,
3. for $i>0$ and $x_{i-1}\in\mathscr{X}_{i-1}$, $p_i(\,.\,|x_{i-1})$ is a conditional probability distribution over $\mathscr{X}_i$.

We call a pair $(\mathscr{X}_i, p_i)_{i=0\ldots,n}$ a *linear causal network.*

From the p_i's we can define the *joint* distributions μ^k on the product space $\mathscr{X}^k := \prod_{i=0}^k \mathscr{X}_i$, $k\leq n$, by

$$\mu^k(x_0,\ldots,x_k) := \prod_{i=0}^{k} p_i(x_i|x_{i-1}). \tag{3.11}$$

We write μ for μ^n. As for each sequence $\mathbf{x}=\langle x_0,\ldots,x_n\rangle\in\mathscr{X}^n$ the probability of x_{i+1} does only depend on its predecessor x_i, the processes defined

by $(\mathscr{X}_i, p_i)_{i=1,\ldots,n}$ has the general properties of a *Markovian* processes (Pearle, 2000, p. 14).

We are now going to introduce the *marginal* probabilities. Let π_i denote the *projection* of $\mathscr{X}^k$ onto $\mathscr{X}_i$; i.e. for $i \leq k$ and $\mathbf{x} = \langle x_0, \ldots, x_k \rangle \in \mathscr{X}^k$ let $\pi_i(\mathbf{x}) := x_i$, and for $X \subseteq \mathscr{X}^k$ let $\pi_i(X) = \{\pi_i(\mathbf{x}) \,|\, \mathbf{x} \in X\}$. For $X \subseteq \mathscr{X}_i$ we set

$$\pi_i^{-1}[X] := \{\mathbf{x} \in \mathscr{X}^n \,|\, \pi_i(\mathbf{x}) \in X\}. \tag{3.12}$$

We define the *marginal* probabilities μ_i on $\mathscr{X}_i$ by:

$$\mu_i(X) = \mu(\pi_i^{-1}[X]), \text{ for } X \subseteq \mathscr{X}_i. \tag{3.13}$$

For $i \leq k \leq n$, $X \subseteq \mathscr{X}_i$, it holds $\mu^k(\pi_i^{-1}[X]) = \mu^n(\pi_i^{-1}[X])$. Hence, the definition of the marginal probabilities μ_i in (3.13) does not depend on the fact that it is defined relative to μ^n. By induction it can be shown that $\mu_i(X)$ equals

$$\sum_{x_0 \in \mathscr{X}_0} p_0(x_0) \sum_{x_1 \in \mathscr{X}_1} p_1(x_1|x_0) \ldots \sum_{x_{i-1} \in \mathscr{X}_{i-1}} p_{i-1}(x_{i-1}|x_{i-2}) \sum_{x_i \in X} p_i(x_i|x_{i-1}) \tag{3.14}$$

Finally, we define *conditional marginal probabilities* $\mu_{i|j}$ as follows: let $X \subseteq \mathscr{X}_i$, and $Y \subseteq \mathscr{X}_j$ with $\mu_j(Y) > 0$, then the conditional marginal probability of X given Y is defined by:

$$\mu_{i|j}(X|Y) = \mu(\pi_i^{-1}[X]|\pi_j^{-1}[Y]). \tag{3.15}$$

With these preparations, we can introduce our general definition of *natural meaning*:

Definition 3.1 *Let $(\mathscr{X}_i, p_i)_{i=0,\ldots,n}$ be a linear causal network. Then, for $X \subseteq \mathscr{X}_i$ and $Y \subseteq \mathscr{X}_j$ with $\mu_j(Y) > 0$, we set*

$$(\mathscr{X}_i, p_i) \models Y \Rrightarrow X :\Longleftrightarrow \mu_{i|j}(X|Y) = 1. \tag{3.16}$$

We say that event Y naturally means *that X*.

If all $\mathscr{X}_i$ are countable, then there is a smallest set X which is naturally implied by the occurrence of an event Y. We can identify this set with the *the natural meaning* of Y.

If X and Y are singletons, i.e. if $X = \{x\}$ and $Y = \{y\}$, then we write $\mu_{i|j}(x|y)$ instead of $\mu_{i|j}(\{x\}|\{y\})$. Furthermore, if i and j are clear from context, e.g. because x can only be an element of $\mathscr{X}_i$, or X a subset of $\mathscr{X}_i$, then we write μ instead of μ_i, or $\mu_{i|j}$.

In (3.16), nothing depends on the fact that $(\mathscr{X}_i, p_i)_{i=0,\ldots,n}$ is a linear causal network. The p_is could equally well depend on any set of random variables $\mathscr{X}_j$ as long as $j < i$. But the condition of linearity plays an important role if we apply the concept of *natural meaning* to signalling games. Here, the fact that

signalling games in the sense of Definition 2.1 define linear causal networks entails that the *common natural information* of speaker and hearer is identical to the hearer's information state! We show this in Lemma 3.4 at the end of this section.

We introduce the relevant notion of *common natural information* in full generality. Let $(\mathscr{X}_i, p_i)_{i=0,\ldots,n}$ be given. For $\mathbf{x} \in \mathscr{X}^n$ and $I \subseteq \{0,\ldots,n\}$ let $\mathbf{x}\restriction_I$ be the restriction of $\mathbf{x}$ to I, i.e. it is the function with domain I and values $(\mathbf{x}\restriction_I)(i) = \pi_i(\mathbf{x})$. We set:

$$[\mathbf{x}\restriction_I] := \{\mathbf{y} \in \mathscr{X}^n \mid \mu(\mathbf{y}) > 0 \wedge \mathbf{x}\restriction_I = \mathbf{y}\restriction_I\}. \tag{3.17}$$

For $\mathbf{x} \in \mathscr{X}^n$ we define the common natural information by the following construction:

$$\begin{aligned} E_{I,J}(\mathbf{x}) &= [\mathbf{x}\restriction_I] \cup [\mathbf{x}\restriction_J], \\ E^0_{I,J}(\mathbf{x}) &= \{\mathbf{x}\}, \\ E^{n+1}_{I,J}(\mathbf{x}) &= \bigcup\{[\mathbf{y}\restriction_I] \cup [\mathbf{y}\restriction_J] \mid \mathbf{y} \in E^n_{I,J}(\mathbf{x})\}, \\ \mathrm{CNI}_{I,J}(\mathbf{x}) &= \bigcup_n E^n_{I,J}(\mathbf{x}). \end{aligned} \tag{3.18}$$

The index sets I and J represent the information states of two agents. Hence, $\mathrm{CNI}_{I,J}(\mathbf{x})$ corresponds to the standard definitions of *common knowledge*. *Implicated* information is generally considered to be part of the common knowledge. As we explicate implicatures as common natural information, we have to spell out what it means that an event Y carries the information that an event X is common natural information. Hence, let $Y \subseteq \mathscr{X}_j$, $X \subseteq \mathscr{X}_i$, and $\mathbf{x} \in \mathscr{X}^n$. We obviously have to conditionalise the conditional marginal probability in (3.16) to $\mathrm{CNI}_{I,J}(\mathbf{x})$; i.e. we have to replace the condition $\mu(\pi_i^{-1}[X]|\pi_j^{-1}[Y]) = 1$ by the condition $\mu(\pi_i^{-1}[X]|\pi_j^{-1}[Y] \cap \mathrm{CNI}_{I,J}(\mathbf{x})) = 1$. First, if this definition should capture the common natural information carried by event Y for two agents represented by the index sets I and J, then Y should be known to both of them, hence, it should hold that $j \in I \cap J$. Second, from this it follows that the condition is reasonable only if $\pi_j(\mathbf{x}) \in Y$. These two restrictions entail that $\mu(\pi_i^{-1}[X]|\pi_j^{-1}[Y] \cap \mathrm{CNI}_{I,J}(\mathbf{x})) = \mu(\pi_i^{-1}[X]|\mathrm{CNI}_{I,J}(\mathbf{x}))$. Hence, the definition of common natural information for a branch $\mathbf{x}$ cannot depend on the set Y of observable values. This straightforwardly leads to the following definition of an event X being common natural information for a branch $\mathbf{x}$ and agents represented by index sets I, J:

Definition 3.2 *Let $(\mathscr{X}_i, p_i)_{i=0,\ldots,n}$ be a linear causal network. Then, for $X \subseteq \mathscr{X}_i$, $\mathbf{x} \in \mathscr{X}^n$ with $\mu(\mathbf{x}) > 0$, we set for $I, J \subseteq \{0,\ldots,n\}$, $I, J \neq \emptyset$:*

$$(\mathscr{X}_i, p_i, \mathbf{x}) \models \mathrm{C}_{I,J}X :\Longleftrightarrow \mu(\pi_i^{-1}[X]|\mathrm{CNI}_{I,J}(\mathbf{x})) = 1. \tag{3.19}$$

We apply these notions to signalling games as follows: For a given signalling game, we identify $\mathscr{X}_0$ with Ω, $\mathscr{X}_1$ with Θ, $\mathscr{X}_2$ with $\mathscr{F}$, and $\mathscr{X}_3$ with $\mathscr{A}$; accordingly, $p_0 = P$, $p_1 = p$, $p_2 = S$, and $p_3 = H$. The information states of the interlocutors are $I = \{1,2\}$ for the speaker and $J = \{2\}$ for the hearer. A branch in the product space $\mathscr{X}^3$ is a sequence $\mathbf{b} = \langle v, \theta, F, a\rangle$. We simplify notation and write $\mathbf{b}(\Omega)$, $\mathbf{b}(\Theta)$, $\mathbf{b}(\mathscr{F})$, and $\mathbf{b}(\mathscr{A})$ instead of $\pi_0(\mathbf{b})$, $\pi_1(\mathbf{b})$, etc.

In signalling games it holds that the hearer's information state J is a subset of the speaker's information state I. This leads to a significant simplification of (3.19). First, we note that it obviously holds that:

$$J \subseteq I \Rightarrow [\mathbf{x}\restriction_I] \subseteq [\mathbf{x}\restriction_J]. \tag{3.20}$$

Furthermore, by induction it can be shown that:

$$i \in I \cap J \Rightarrow \forall n > 0 \forall \mathbf{y} \in E^n_{I,J}(\mathbf{x})\, \pi_i(\mathbf{y}) = \pi_i(\mathbf{x}). \tag{3.21}$$

From these two facts, it follows by induction that $J \subseteq I$ implies that $\forall n > 0 E^n_{I,J}(\mathbf{x}) = [\mathbf{x}\restriction_J]$, and hence that:

$$J \subseteq I \Rightarrow \mathrm{CNI}_{I,J}(\mathbf{x}) = [\mathbf{x}\restriction_J]. \tag{3.22}$$

Identifying *implicatures* of an utterance F with the common natural information carried by this event, we arrive at:

Definition 3.3 (Implicature) *Let (S,H) be a strategy pair for a signalling game $\mathscr{G} = \langle \Omega, \Theta, P, p, \mathscr{F}, \mathscr{A}, u\rangle$. Let $(\mathscr{X}_i, p_i)_{i=0,\ldots,n}$ be the linear causal network defined by identifying $\mathscr{X}_0$ with Ω, $\mathscr{X}_1$ with Θ, $\mathscr{X}_2$ with $\mathscr{F}$, and $\mathscr{X}_3$ with $\mathscr{A}$; accordingly, $p_0 = P$, $p_1 = p$, $p_2 = S$, and $p_3 = H$. Let $X \subseteq \mathscr{X}_i$, $I = \{1,2\}$ and $J = \{2\}$. Let μ be the probability distribution on the product space $\mathscr{X}^3$ defined in (3.11), and let $\mathbf{b}$ be a branch in $\mathscr{X}^3$ with $\mu(\mathbf{b}) > 0$. Then we set for $\mathbf{b}(\mathscr{F}) = F$:*

$$\langle \mathscr{G}, S, H, \mathbf{b}\rangle \models F +\!> X :\Longleftrightarrow\ (\mathscr{X}_i, p_i, \mathbf{b}) \models \mathrm{C}_{I,J} X. \tag{3.23}$$

We then say that in $\mathbf{b}$ the utterance of F implicates *that X. We simply say that the utterance of F* implicates *that X, $\langle \mathscr{G}, S, H\rangle \models Y +\!> X$, if $\langle \mathscr{G}, S, H, \mathbf{b}\rangle \models F +\!> X$ for all $\mathbf{b}$ for which $\mathbf{b}(\mathscr{F}) = F$ and $\mu(\mathbf{b}) > 0$. Then, for $Y \subseteq \mathscr{F}$, we generalise:*

$$\langle \mathscr{G}, S, H\rangle \models Y +\!> X :\Longleftrightarrow\ \forall F \in Y\ \langle \mathscr{G}, S, H\rangle \models F +\!> X. \tag{3.24}$$

According to the generalisation in (3.24), a set Y of signals implicates X if *every* form $F \in Y$ implicates X. By (3.22), it immediately follows that:

Lemma 3.4 *Let $\mathscr{G} = \langle \Omega, \Theta, P, p, \mathscr{F}, \mathscr{A}, u\rangle$ be a signalling game, and (S,H) a strategy pair for $\mathscr{G}$. Let $\mu_{i|\mathscr{F}} := \mu_{i|2}$ be the conditional marginal probability distribution defined in (3.15) for the linear causal network $(\mathscr{X}_i, p_i)_{i=0,\ldots,3}$ defined*

by $\langle \mathscr{G}, S, H \rangle$*. Then, for* $X \subseteq \mathscr{X}_i$*,* $Y \subseteq \mathscr{F}$*, it holds:*

$$\langle \mathscr{G}, S, H \rangle \models Y +\!> X \iff \mu_{i|\mathscr{F}}(X|Y) = 1 \tag{3.25}$$

In the following, we will often identify a solved signalling game $\langle \mathscr{G}, S, H \rangle$ with its associated linear causal network $(\mathscr{X}_i, p_i)_{i=0,\ldots,3}$ and write e.g. $\langle \mathscr{G}, S, H \rangle \models Y \Rrightarrow X$ iff $(\mathscr{X}_i, p_i)_{i=0,\ldots,3} \models Y \Rrightarrow X$ in the sense of Def. 3.1. Using this convention, we can rewrite (3.25) equivalently as

$$\langle \mathscr{G}, S, H \rangle \models Y +\!> X \iff \langle \mathscr{G}, S, H \rangle \models Y \Rrightarrow X, \tag{3.26}$$

i.e. *Y implicates X* iff *Y naturally means X*.

We further explore the potential of Definition 3.3 in Section 5.

4 The Solution Concept

4.1 Preliminary Remarks

With the terminology of Section 3, the conditions of Lemma 2.3 can now be reformulated as follows: If $\langle \mathscr{G}, S, H \rangle$ is such that an utterance of F *naturally means* that the hearer chooses a speaker optimal act, then (S, H) is a Pareto Nash equilibrium; if $\langle \mathscr{G}, S, H \rangle$ is such that an utterance of F does *not* naturally mean that the hearer chooses a speaker optimal act, then (S, H) is *not* a Pareto Nash equilibrium. We mentioned before that we interpret the probabilities in signalling games as objective probabilities. Hence, Lemma 2.3 provides us with a criterion for deciding whether a strategy pair is an *objective* Pareto Nash equilibrium.

In principle, there are two interpretations of probabilities which are of interest to us: the interpretation as objective frequencies, and the interpretation as subjective probabilities in the sense of (Savage, 1972). We will use both interpretations depending on which aspect of communication we are modelling. We interpret probabilities objectively if we want to explain the objective success of communication seen as a real world phenomenon; we interpret them subjectively if we model the cognitive level. Objective probabilities are just the familiar relative frequencies. Subjective probabilities are mathematical constructs which offer concise representations of the agent's propensities for choosing actions; i.e. assigning subjective probability P_X and utility function u_X to agent X means that X's preferences over actions a after learning F are indistinguishable from an agent's preferences who chooses between actions according to the expected utilities $EU_X(a|F)$. As subjective probabilities are mathematical constructs, assigning them to agents does not mean that these agents actually represent these probabilities, or reason with them. Likewise, subjective probabilities do, in general, not have to correspond to observable frequencies. Objective frequencies may be completely unknown to our interlocutors; it may even

be that they don't even possess a notion of *probability*. As the probabilities P and p defined in signalling games represent the probabilities with which *nature* is choosing worlds and speaker's types, they have to be interpreted as objective frequencies, hence they might not be known to the interlocutors. In this section, we provide a model of the communicative situation which only represents the interlocutors' subjective expectations about the state of the world but not the objective frequencies with which nature chooses the world or the speaker's type.

The task is to describe the communicative situation in terms of its cognitively relevant parameters, and to provide a method for finding solutions (S,H) to the coordination problem posed by the communicative situation. As our models are intended as models of online communication, it is prima facia reasonable to look for a method which is *as simple as possible*.

In most game theoretic models, equilibrium concepts are describing the stable patterns of behaviour which can emerge from the interaction of rational agents in certain classes of games. As different populations playing these games may adopt different behaviours, the task in empirical applications is to find the set of all possible strategy profiles which satisfy a given equilibrium concept and to show that the behavioural patterns found in the different populations correspond to one or the other strategy profile in this set. In this paper, we follow a different strategy. We assume that there is a signalling strategy established in the population which defines the *semantic* meaning of signals (Lewis, 2002); i.e. we assume that the speaker's signals have a predefined meaning which restricts their use. The pure semantic meaning of signals also defines a hearer strategy for choosing between available actions after learning the signal's semantic meaning. Starting out from this situation, we are interested in the Nash equilibrium (S,H) which is *closest* to the given semantic convention. We think of the *distance* in terms of the number of steps of reasoning about each other which are involved in reaching the equilibrium. This can be made more precise in the framework of *iterated best response* (IBR) models (Jäger and Ebert, 2009; Franke, 2009).[1] IBR models explicate the reasoning about each other by an iterated process. In each step of this process, one of the two interlocutors chooses a best response strategy to the strategy which he assumes the other interlocutor has chosen in the previous step. There are two possible strategies from which the IBR process can start: the process can either start with a speaker strategy or with a hearer strategy. Accordingly, the model consists of two separate lines of reasoning. These two lines are shown in Figure 3.

In the IBR models worked out by (Jäger and Ebert, 2009; Franke, 2009), the S_i and R_i are in fact *sets* of strategies. In (Franke, 2009), S_0 is the set of

[1]The following sketch of the IBR model is a simplified version of (Franke, 2009). For more details, motivation, and differences between the models, we refer to the original papers.

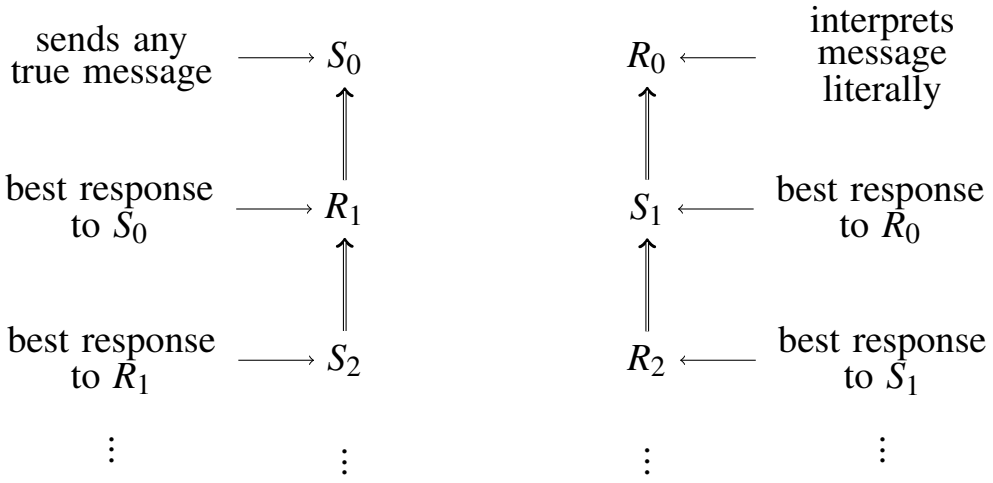


Fig. 3: Schema of the IBR-sequence (Franke, 2009, p. 57).

all speaker strategies for which the speaker arbitrarily chooses a signal which he believes to be true. Hence, the S_0–speakers do not take the hearer's strategy into account. The hearer chooses an action after receiving the speaker's signal. Receiving it, he learns the semantic content of it. R_0 is the set of all hearer strategies for which the hearer only takes the semantic meaning of signals into account. Hence, R_0–hearers do not reason about the speaker. This means that on the 0–level it suffices to know the shared utilities and the speaker's and hearer's (subjective) probabilities about the state of affairs for defining S_0 and R_0. In step $n+1$ of the IBR process, each interlocutor I assumes that the other interlocutor J adopts a certain strategy from J's strategy set defined in the nth step. Together with I's expectations about the state of affairs, this defines I's new set of best response strategies. This means, e.g., that, in the first step from S_0 to R_1, the hearer assumes that the speaker adopted some S_0 strategy, which arbitrarily chooses a sentence which the speaker believes to be true. The hearer, after receiving a signal F, then chooses an act which has the highest expected utility given the fact that the speaker sent F. R_1 is then the set of all hearer strategies which, in this way, can result as a best response to some $S \in S_0$. Similarly, in the first iteration step from R_0 to S_1, the speaker assumes that the hearer follows some strategy in R_0. The speaker, as a response, chooses signals which lead the hearer to choose such actions which will have the highest expected utility as seen from the speaker's perspective. This defines the set S_1. This process can be iterated. IBR models then look for pairs of strategy sets (S^*, H^*) which eventually become *stable*.[2]

How many iteration steps does it at least take to reach a stable state? We can consider the two lines of the IBR model separately as strategy sets occurring

[2]Stability is defined by a *looping* condition for the strategy sets S^* and R^*. For details, see (Franke, 2009, p. 58).

in one line have no influence on the strategy sets in the other line. Hence, let us consider the line starting with the speaker strategies in S_0. The hearers set of best responses R_1 will in general be different from R_0 as the fact that a signal was sent may carry information in addition to the semantic meaning of the signal. As the strategies in S_0 randomly produced true signals, S_2, the speaker's best responses to R_1, will in general be different to S_0. Hence, a stable state cannot be reached before S_2 is reached. The earliest stage at which the hearer can see that he has reached a stable state is therefore the stage in which he calculates R_3; and the earliest stage at which the speaker can see that he has reached a stable state is, accordingly, the stage in which he calculates S_4. Hence, for the line starting with S_0, for reaching a stable state, the hearer must at least consider the speaker's best response to his best response to the speaker's random strategy; and the speaker has at least to consider the hearer's best responses to the speaker's best responses to the hearer's best responses to the speaker's random strategies in S_0. Let us now turn to the line of the IBR model starting with R_0. The earliest stage at which the hearer can see that he has reached a stable state is the stage in which he calculates R_2; and the earliest stage at which the speaker can see that he has reached a stable state is, accordingly, the stage in which he calculates S_3. Hence, for the line starting with R_0, the hearer must at least consider the speaker's best response to his basic strategies in R_0, and the speaker has at least to consider the hearer's best responses to the speaker's best responses to the hearer's basic strategies. As R_0 is, in general, not identical to R_1, the speaker's set S_1 of best responses to R_0 will, in general, also be different from S_2. Hence, if one line stops at an early stage, it is no guarantee that the other line does also stop early. If we take the IBR model serious as a cognitive model, then these reasoning steps must be a cognitive reality. In this section, we show that the coordination problem posed by communication can be solved with fewer steps of reasoning about each other than predicted by the IBR model. More precisely, we show that backward induction provides a solution which guarantees that speaker and hearer have reached a stable strategy pair without having to calculate *whether they have reached a stable state*.

The IBR model shows that, in order to find out whether a strategy is stable by reasoning about each other, the hearer must take into account the speaker's best response to a hearer strategy at least once. Hence, the shortest possible path to a stable strategy is the R_0–S_1–R_2–S_3–path. If the method for finding a stable solution should be *simpler* or *shorter* than the method provided by the IBR model, then we have to find a method which avoids some steps of reasoning about each other in this sequence. In this respect, the simplest method is backward induction. When applying backward induction to a signalling game $\mathscr{G}$, the hearer does never consider the speaker's strategy, and the speaker considers the hearer's strategy only once. This is the cognitively least demanding method

for finding solutions. We will show in Section 4.3 that the resulting strategy pair (S,H) guarantees that for any possible utterance the signal *naturally means* that the hearer chooses a speaker optimal act. From Lemma 2.3 it follows that (S,H) is a Pareto Nash equilibrium; hence it is a stable strategy pair. There is no need for further steps of reasoning about each other. The following method for finding a solution to the coordination problem described by signalling games was introduced in (Benz, 2006). We call it *the Optimal–Answer* (OA) model.

4.2 The Optimal–Answer Model

In this section, the general features of the communicative situation are the same as that considered in the context of signalling games. We again assume that the conversation is subordinated to a joint purpose which is defined by a decision problem of the hearer. This decision problem may be revealed by an implicit or explicit question by the hearer. Hence, we can call the speaker's message an *answer*. The OA model tells us which answer a rational language user will choose given the hearer's decision problem and his knowledge about the world. We call the basic models which represent the utterance situation as *support problems*. They consist of the hearer's decision problem and the speaker's expectations about the world. These expectations are represented by subjective probabilities. In (Benz, 2006, 2007), it was shown that, in general, it is not possible to define a reliable *relevance* measure such that the speaker may simply maximise the relevance of his answers for optimally supporting the hearer. When solving a support problem the speaker has to take the hearer's response to his choice of signal into account. Hence, in view of our previous discussion of IBR models, this shows that there is no reliable method of solving a support problem which involves fewer steps of reasoning about each other than backward induction. Support problems incorporate Grice' *Cooperative Principle*, his maxim of *Quality*, and a method for finding optimal strategies which replaces Grice' maxims of *Quantity* and *Relevance*. For now, 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 $\mathscr{A}$ 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 $F \subseteq \Omega$ it is $P(F) = \sum_{v\in F} P(v)$. 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 world–action pairs. We collect these elements in the following structure:

Definition 4.1 *A* decision problem *is a triple* $\langle(\Omega,P),\mathscr{A},u\rangle$ *such that* (Ω,P) *is a finite probability space,* $\mathscr{A}$ *a finite, non–empty set and* $u: \Omega\times\mathscr{A} \to \mathbb{R}$

a function. $\mathscr{A}$ *is called the* action *set, and its elements* actions*;* u *is called a* payoff *or* utility *function.*

In the following, a decision problem $\langle(\Omega,P),\mathscr{A},u\rangle$ represents the hearer'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. In Section 2, we used the symbol $\mathscr{E}$ to denote the expected utility. As in the present section probabilities are assumed to be subjective probabilities, we use different notation in order to distinguish subjective expected utilities from expected utilities defined from objective frequencies. Hence, we write for the (subjective) *expected utility* of action $a \in \mathscr{A}$ in decision problem $\langle(\Omega,P),\mathscr{A},u\rangle$:

$$EU(a) = \sum_{v\in\Omega} P(v) \times u(v,a). \tag{4.27}$$

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 hearer's beliefs. In probability theory the result of learning a proposition F is modelled by *conditional probabilities*. Let H be any proposition and F the newly learned proposition. Then, the probability of H *given* F, written $P(H|F)$, is defined as

$$P(H|F) := P(H\cap A)/P(F) \text{ for } P(F) \neq 0. \tag{4.28}$$

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

$$EU(a|F) = \sum_{v\in\Omega} P(v|F) \times u(v,a). \tag{4.29}$$

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

$$\mathscr{B}(F) := \{a \in \mathscr{A} \mid \forall b \in \mathscr{A}\ EU_H(b|F) \leq EU_H(a|F)\}. \tag{4.30}$$

The decision problem represents the hearer's situation. In order to get a model of the questioning and answering situation, we have to add a representation of the answering speaker's information state. We identify it with a (subjective) probability distribution P_S that represents his expectations about the world. We make a number of assumptions in order to match the definition of support problems to our previous definition of signalling games. First, we assume that

the hearer's expectations are common knowledge. Second, we assume that there exists a common prior from which both the speaker's and the hearer's information state can be derived by a Bayesian update. This entails that the speakers and the hearer's expectations cannot contradict each other. Third, we assume that the speaker does not directly choose propositions but linguistic *forms* or *signals* which have a predefined semantics. Furthermore, we assume that the forms $F \in \mathscr{F}$ come with positive costs. This leads to the following definition of *interpreted* support problems:

Definition 4.2 *A tuple* $\sigma = \langle \Omega, P_S, P_H, \mathscr{F}, \mathscr{A}, u, c, [\![.]\!] \rangle$ *is an* interpreted support problem *if:*

1. (Ω, P_S) *is a finite probability space and* $\langle (\Omega, P_H), \mathscr{A}, u \rangle$ *a decision problem;*
2. *there exists a probability distribution* P *on* Ω*, and sets* $K_S \subseteq K_H \subseteq \Omega$ *for which* $P_S(X) = P(X|K_S)$ *and* $P_H(X) = P(X|K_H)$*;*
3. $[\![.]\!] : \mathscr{F} \to \mathscr{P}(\Omega)$ *is an interpretation function for the elements* $F \in \mathscr{F}$*. We assume that*

$$\forall X \subseteq \Omega \, \exists F \in \mathscr{F} \; [\![F]\!] = X; \tag{4.31}$$

4. $u : \Omega \times \mathscr{A} \to \mathbb{R}$ *is a utility measure and* c *a cost function that maps forms* $F \in \mathscr{F}$ *to positive real number.*

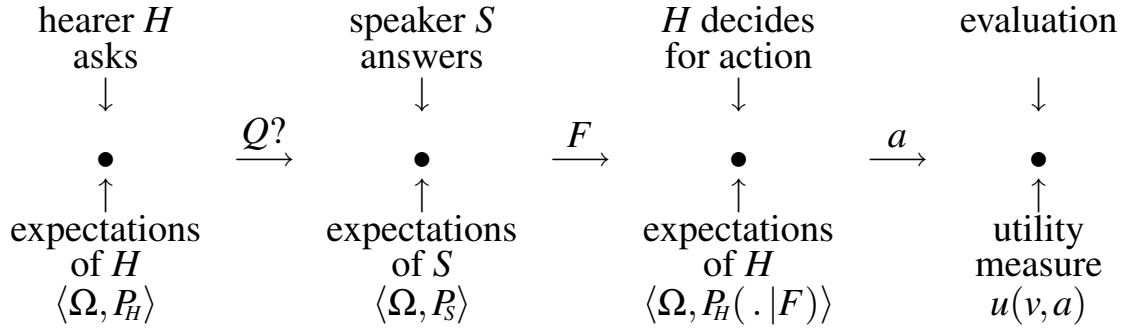
The second condition says that P_S and P_H are derived from a common prior P by a Bayesian update. It entails:

$$\forall X \subseteq \Omega \; P_S(X) = P_H(X|K_S). \tag{4.32}$$

This condition allows us to identify the *common ground* in conversation with the addressee's expectations about the domain Ω, i.e. with P_H. 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 and Stockhof, 1991). Furthermore, condition (4.32) implies that the speaker's beliefs cannot contradict the hearer's expectations, i.e. for $X \subseteq \Omega$: $P_S(X) = 1 \Rightarrow P_H(X) > 0$.

In order to simplify notation, we will often write F instead of $[\![F]\!]$. Hence, F may denote a proposition or a linguistic form, depending on context.

Our next goal is to introduce a principle for solving support problems, i.e. for finding the speaker's and hearer's strategies which lead to optimal outcomes. The speaker 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:



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 *Co-operative 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 $a \in \mathscr{B}(F)$ 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 $h(.|F)$ on $\mathscr{A}$. Then, the expected utility of an answer F is defined by:

$$EU_S(F) := \sum_{a \in \mathscr{B}(F)} h(a|F) \times EU_S(a). \tag{4.33}$$

We add here a further Gricean maxim, the *Maxim of Quality*. We call an answer F *admissible* if $P_S(F) = 1$. The Maxim of Quality is represented by the assumption that the speaker S does only give admissible answers. This means that he believes them to be *true*. For an interpreted support problem $\sigma = \langle \Omega, P_S, P_H, \mathscr{F}, \mathscr{A}, u, c, [\![\,.\,]\!] \rangle$ we set:

$$\mathrm{Adm}_\sigma := \{F \subseteq \Omega \,|\, P_S(F) = 1\} \tag{4.34}$$

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

$$\mathrm{Op}_\sigma := \{F \in \mathrm{Adm}_\sigma \,|\, \forall B \in \mathrm{Adm}_\sigma\, EU_S(B) \leq EU_S(F)\}. \tag{4.35}$$

We write Op^h_σ 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. Condition (4.31), it follows that all propositions $A \subseteq \Omega$ can be expressed. Hence, we can think of Op_σ as a subset of $\mathscr{P}(\Omega)$ or as a subset of $\mathscr{F}$.

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 $h(.|F)$ is an example of a mixed strategy. We define a (mixed) strategy pair for an interpreted support problem σ to be a pair (s,h) such that s is a probability distribution over $\mathscr{F}$ and $h(.|F)$ a probability distribution over $\mathscr{A}$.

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

Definition 4.3 *Let $\sigma = \langle \Omega, P_S, P_H, \mathscr{F}, \mathscr{A}, u, c, [\![\, . \,]\!] \rangle$ be a given interpreted support problem. The* canonical solution *to σ is a pair (S,H) of mixed strategies which satisfy:*

$$S(F) = \begin{cases} |\mathrm{Op}_\sigma|^{-1}, F \in \mathrm{Op}_\sigma \\ 0 \textit{ otherwise} \end{cases}, H(a|F) = \begin{cases} |\mathscr{B}(F)|^{-1}, a \in \mathscr{B}(F) \\ 0 \textit{ otherwise} \end{cases}. \quad (4.36)$$

We write $S(\,.\,|\sigma)$ if S is a function that maps each $\sigma \in \mathscr{S}$ 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 $H(a|F) > 0 \Leftrightarrow a \in \mathscr{B}(F)$ and $S(F|\sigma) > 0 \Leftrightarrow F \in \mathrm{Op}_\sigma$.

The speaker may always answer everything he knows, i.e. he may answer $K_S := \{v \in \Omega \,|\, P_S(v) > 0\}$. Condition (4.32) trivially entails that $\mathscr{B}(K_S) = \{a \in \mathscr{A} \,|\, \forall b \in \mathscr{A}\ EU_S(b) \leq EU_S(a)\}$. If speaker and hearer follow the canonical solution, and if we ignore the different costs of answers, then:

$$\mathrm{Op}_\sigma = \{F \in \mathrm{Adm}_\sigma \,|\, \mathscr{B}(F) \subseteq \mathscr{B}(K_S)\}. \quad (4.37)$$

In order to show (4.37), let $F \in \mathrm{Adm}$ and $\alpha := \max\{EU_S(a) \,|\, a \in \mathscr{A}\}$. For $a \in \mathscr{B}(F) \setminus \mathscr{B}(K_S)$ it holds by definition that $EU_S(a) < \alpha$ and $H(a|F) > 0$. $EU_S(F)$ is the sum of all $H(a|F) \times EU_S(a)$. If $\mathscr{B}(F) \not\subseteq \mathscr{B}(K_S)$, then this sum divides into the sum over all $a \in \mathscr{B}(F) \setminus \mathscr{B}(K_S)$ and all $a \in \mathscr{B}(F) \cap \mathscr{B}(K_S)$. Hence, $EU_S(F) < \alpha$, and therefore $F \notin \mathrm{Op}_\sigma$.

If $\mathscr{B}(F) \not\subseteq \mathscr{B}(K_S)$, then the speaker knows that answering F would induce the addressee to choose a sub–optimal action with positive probability. In this sense, we can call an answer F *misleading* if $\mathscr{B}(F) \not\subseteq \mathscr{B}(K_S)$; then, (4.37) implies that Op_σ is the set of all non–misleading answers.

4.3 Signalling Games and the Optimal Answer Model

We first recall the definition of signalling games from the previous sections. A *signalling game* is a tuple $\langle \Omega, \Theta, P, p, \mathscr{F}, \mathscr{A}, u \rangle$ for which: (1) Ω and Θ are non–empty finite sets; (2) $P(\,.\,)$ is a probability distribution over Ω; (3) $p(\,.\,|v)$ is a probability distribution over Θ for every $v \in \Omega$; (4) $\mathscr{F}$ and $\mathscr{A}$ are respectively

the speaker's and hearer's action sets; and (5) $u : \Omega \times \Theta \times \mathscr{F} \times \mathscr{A} \to \mathbb{R}$ is a shared utility function. We also assumed that $u(v, \theta, F, a)$ can be decomposed into $u(v,a) - c(F)$ for some positive value $c(F)$.

We first discuss the consequences of interpreting the probabilities for signalling games as objective frequencies and that for support problems as subjective probabilities.

If $\mathscr{S}$ is a set of support problems with identical decision problems, we can construct a corresponding signalling game. As it is assumed that the speaker knows the full support problem, we can identify $\mathscr{S}$ with the set of speaker's types Θ. The action sets and the utility function of the signalling game are just the same as that of the support problems. As the decision problems of the support problems in $\mathscr{S}$ are identical, this poses no problem. The only non–trivial correspondence is that of the probabilities.

As mentioned before, we regard the probabilities P and p of the signalling game as objective frequencies. Under this interpretation, Lemma 2.3 states the objective conditions for optimal signalling strategies. If we interpret P_S^σ and P_H^σ as the agents' representations for these objective probabilities, then P_S must be identical to μ_S, and P_H to P.[3] (4.32) then entails that $P_H(v|K_S) = \mu_S(v|\sigma)$. It holds $P_H(v|K_S) = \mu_S(v|\sigma)$ *iff* $P(v)/P(K_S) = P(v)\, p(\sigma|v)/\mu(\sigma)$ *iff* $p(\sigma|v) = \mu(\sigma)/P(K_S)$. The last term does not depend on v, hence, it follows that (4.32) entails that $p(\sigma|v)$ must be the same for all $v \in K_S$.

In (Benz and van Rooij, 2007), we identified P_S with $P(\,.\,|K_S)$, and P_H with P. Then (4.32) trivially holds. p was considered to be a representation of the hearer's subjective expectations about the speaker's types. In order to distinguish the hearer's subjective probabilities about the speaker's type from the objective frequencies, we write p_H for the former, and keep p for the latter. Subjective probabilities per se have no causal influence on the objective probabilities. Hence, p_H is logically independent from P and p. Under this interpretation, it can be shown that the strategy pair (S,H) defined by the canonical solutions to the support problems (4.36) is optimal for all possible p_H. This result follows from Lemma 2.3 if we assume that the objective frequencies represented by p in the signalling game again satisfy $p(\sigma|v) = \mu(\sigma)/P(K_S)$. Then, whatever the subjective expectations of the hearer about the speaker's types are, the canonical strategy will satisfy (2.9), and hence be optimal in the sense that there is no other strategy pair with higher expected utility.

In this paper, we go one step further and completely separate the subjective cognitive level from the objective level. Hence, we interpret the probabilities P_S and P_H in the support problems as subjective probabilities which are logically independent of the frequencies P and p of the underlying signalling game. As P_S and P_H are subjective, they don't change the objective information

[3]The probabilities μ_S and μ have been defined in (2.2) and (2.3).

available to S and H. Hence, we can freely assign these probabilities to the interlocutors without changing the signalling game on the objective level. Subjective probabilities determine the speaker's and the hearer's strategies. These strategies are the only connection between the cognitive and the realistic level.

What is the advantage of separating the cognitive and the objective level? There are two issues involved: the *epistemic* issue of the recognisability of objective frequencies, and the issue of *bounded rationality*. For the epistemic issue, the objective frequencies are largely unknown to the interlocutors. The speaker may learn his type θ e.g. by direct observation, by an inductive inference, by hear-say, or from a conversation with someone else. Hence, there are so many and so varied sources for the acquisition of belief type θ that it is not to be expected that the hearer or the speaker can provide any justified estimate of $p(\theta|v)$. In this respect, conversation can be characterised as a game of *complete uncertainty*. Even though, we can assign rationally justified subjective probabilities which describe the agent's behaviour on the cognitive level. This move allows us to treat communication as a game under *risk*. For the issue of bounded rationality, it doesn't deem us a realistic assumption that interlocutors do an online calculation of their conditional probabilities μ_S and μ_H defined in (2.2). The established solution concept for signalling games is that of a perfect Baysian equilibrium. Hence, even if we could assume that the interlocutors know the objective frequencies P and p, the complexity of calculating the Bayesian perfect equilibria would make the resulting model cognitively implausible. By separating the cognitive and the objective level of reality, we can justify simpler solutions to the coordination problem, and at the same time explain their objective success.

What is our approach to the problem of bounded rationality? If we want to show that a strategy pair (S,H) is a successful solution to a signalling game, we have to show that it is a Perfect Baysian equilibrium in the objective sense. We will even show that the strategies established on the cognitive level are such that they Pareto dominate all other solutions. Hence, our strategy for solving the problem of bounded rationality is to search for the simplest solution on the cognitive level that can guarantees objective success. As the discussion of relevance scale approaches in (Benz, 2006, 2007) shows, the interlocutors have to solve a game theoretic problem, i.e. it is not possible to guarantee objective communicative success by simply applying decision theoretically defined solutions on the cognitive level. Signalling games are sequential games. The simplest solution to a sequential game is that found by backward induction. Hence, the optimal answer model claims that the most simple solution concept for sequential games is already successful. Moreover, it involves that the hearer does not need to take his expectations p_H about the speaker's types θ into account. This leads to our main criterion of simplicity: we assume that a method for finding

a solution (S,H) is the simpler the less reasoning about each other is involved in it. In terms of the IBR model, this means that a R_0–S_1 reasoning sequence is sufficient for finding reliable stable equilibria.

In order to decide whether the canonical strategy determined by a set of support problems is a Pareto optimal equilibrium for the related signalling game, the logical relation between the objective frequencies of signalling games and the subjective probabilities of sets of support problems play a central role. We consider the following relations:

Definition 4.4 *Let $\mathscr{S}$ be a set of interpreted support problems. Let's assume that the support problems $\sigma = \langle \Omega, P_S, P_H, \mathscr{F}, \mathscr{A}, u, c, [\![.]\!] \rangle$ may only differ with respect to P_S^σ. Let $\mathscr{G} = \langle \Omega, \Theta, P, p, \mathscr{F}, \mathscr{A}, u \rangle$ be any basic signalling game for which $\Theta = \mathscr{S}$ and $\mu_\Theta(\sigma) = \sum_v P(v)\, p(\sigma|v) > 0$ for all $\sigma \in \mathscr{S}$. We call the speaker's probability P_S^σ:*

1. fully reliable *if* $P_S^\sigma = \mu_S(\,.\,|\sigma)$.
2. reliable *if* $\forall v \in \Omega\, (\mu_S(v|\sigma) > 0 \Leftrightarrow P_S^\sigma(v) > 0)$.
3. truth preserving *if* $\forall v \in \Omega\, (\mu_S(v|\sigma) > 0 \Rightarrow P_S^\sigma(v) > 0)$.

We say that:

4. $\mathscr{G}$ supports $\mathscr{S}$ *iff all P_S^σ are reliable;*
5. $\mathscr{G}$ fully *supports $\mathscr{S}$ iff all P_S^σ are fully reliable;*
6. $\mathscr{G}$ weakly *supports $\mathscr{S}$ iff all P_S^σ are truth preserving.*

Full reliability is stronger than reliability, and reliability is stronger than truth preservingness. If P_S is truth preserving then all believes of S are true in the sense that $P_S^\sigma(F) = 1$ implies that the true state of the world must be an element of F. This follows from $P(v) = 0 \Rightarrow \mu_S(v|\sigma) = P(v)\, p(\sigma|v) = 0$.

Furthermore, we introduce two conventions: (1) If the support problem does not specify a set of utterances $\mathscr{F}$ or costs of signals, then we assume that for supporting signalling games it holds that $\mathscr{F} = \mathscr{P}(\Omega)$, and that $u(v,\theta,F,a)$ does only depend on v and a. (2) We also use the terminology of Def. 4.4 if Θ and $\mathscr{S}$ can only be identified with each other by a bijective map. In this case, we write θ_σ and σ_θ for the speaker type and the support problem which have been identified with each other.

The following two lemmas provide the justification for the optimal answer approach. The first one tells us that the canonical solution to a set of support problems is a Pareto Nash equilibrium for all fully supporting signalling games. The second lemma strengthens this result for support problems with expert speaker. In this case, the canonical solution is a Pareto Nash equilibrium to all weakly supporting signalling games.

Lemma 4.5 *Let $\mathscr{S}$ be a set of interpreted support problems. Let's assume that the support problems $\sigma = \langle \Omega, P_S, P_H, \mathscr{F}, \mathscr{A}, u, c, [\![.]\!] \rangle$ may only differ with respect to P_S^σ. Let (S,H) be the canonical solution to $\mathscr{S}$. Let $\mathscr{G} = \langle \Omega, \Theta, P, p, \mathscr{F}, \mathscr{A}, u \rangle$ be any basic signalling game which fully supports $\mathscr{S}$, i.e. $\Theta = \mathscr{S}$ and the speaker's probabilities P_S^σ are* fully reliable. *Then (S,H) is a Pareto Nash equilibrium of $\mathscr{G}$.*

Proof: The lemma follows if we can show that the canonical solution satisfies (2.9) for all $F \in \mathscr{F}$. Hence, let F be given, and σ be such that $\exists v\, P(v)\, p(\sigma|v) > 0$. By definition, $S(F|\sigma) > 0$ iff $F \in \mathrm{Op}_\sigma$; hence, it follows from (4.37) and the definition of the canonical hearer strategy that $H(a|F) > 0$ entails $a \in \mathscr{B}(K_S^\sigma)$ with $K_S^\sigma = \{v \in \Omega \,|\, P_S^\sigma(v) > 0\}$. As P_S is fully reliable, it follows that $\mathscr{B}(K_S^\sigma) = \mathscr{B}(\sigma)$, and therefore that $H(a|F) > 0 \Rightarrow a \in \mathscr{B}(\sigma)$. Hence, $S(F|\sigma) > 0 \Rightarrow H(\mathscr{B}(\sigma)|F) = 1$. ■

For support problems with expert speakers, we arrive at a stronger result:

Lemma 4.6 *Let $\mathscr{S}$ be a set of interpreted support problems. Let's assume that the support problems $\sigma = \langle \Omega, P_S, P_H, \mathscr{F}, \mathscr{A}, u, c, [\![.]\!] \rangle$ may only differ with respect to P_S^σ. Let us further assume that the speaker is an* expert, *i.e.*

$$\forall \sigma \in \mathscr{S} \exists a \in \mathscr{A}\; P_S^\sigma(O(a)) = 1.$$

Let (S,H) be the canonical solution to $\mathscr{S}$. Let $\mathscr{G} = \langle \Omega, \Theta, P, p, \mathscr{F}, \mathscr{A}, u \rangle$ be any signalling game which weakly supports S. Then (S,H) is a Pareto Nash equilibrium of $\mathscr{G}$.

Proof: That the speaker is an expert entails that $\mathscr{B}(K_S^\sigma) = \{a \in \mathscr{A} \,|\, P_S^\sigma(O(a)) = 1\}$. As $\mu_S(v|\sigma) > 0 \Rightarrow P_S^\sigma(v) > 0$, it follows that $\mathscr{B}(K_S^\sigma) \subseteq \mathscr{B}(\sigma)$. Hence, the claim follows as in the proof of Lemma 4.5. ■

It is an obvious question, how to construct a signalling game $\mathscr{G}$ for a given set of support problems $\mathscr{S}$ so that $\mathscr{G}$ is fully supporting $\mathscr{S}$. The answer will be provided by the next lemma. Finally, we will also address the question how and when we can construct a set $\mathscr{S}$ of support problems for a given signalling game $\mathscr{G}$ such that $\mathscr{G}$ supports $\mathscr{S}$.

Lemma 4.7 *Let $\mathscr{S}$ be a set of interpreted support problems. Let's assume that the support problems $\sigma = \langle \Omega, P_S, P_H, \mathscr{F}, \mathscr{A}, u, c, [\![.]\!] \rangle$ may only differ with respect to P_S^σ. Let μ be any probability measure on $\mathscr{S}$ for which $\mu(\sigma) > 0$ for all $\sigma \in \mathscr{S}$. Then let $\nu(v,\sigma) := \mu(\sigma)\, P_S^\sigma(v)$, $P(v) := \sum_\sigma \nu(v,\sigma)$, and $p(\sigma|v) := \nu(v,\sigma)/P(v)$. Then ν is a probability measure on $\Omega \times \mathscr{S}$, and $\mathscr{G} = \langle \Omega, \Theta, P, p, \mathscr{F}, \mathscr{A}, u \rangle$ is fully supporting $\mathscr{S}$.*

Proof: As $\sum_{v,\sigma} \mu(\sigma)\, P_S^\sigma(v) = \sum_\sigma \mu(\sigma) \sum_v P_S^\sigma(v) = 1$, ν is a probability measure on $\Omega \times \mathscr{S}$. That $\mathscr{G}$ supports $\mathscr{S}$ follows from $\mu_\Theta(\sigma) = \sum_w P(w)\, p(\sigma|w) =$

$\sum_w \nu(w,\sigma) = \mu(\sigma) \sum_w P_S^\sigma(w) = \mu(\sigma)$; hence $\mu_\Theta(\sigma) > 0$ for all $\sigma \in \mathscr{S}$. Finally, $\mu_S(v|\sigma) = \frac{P(v)\,p(\sigma|v)}{\sum_w P(w)\,p(\sigma|w)} = \frac{\nu(v,\sigma)}{\sum_w \nu(w,\sigma)} = \frac{P_S^\sigma(v)}{\sum_w P_S^\sigma(w)} = P_S^\sigma(v)$. Hence, $\mathscr{G} = \langle \Omega, \Theta, P, p, \mathscr{F}, \mathscr{A}, u \rangle$ is fully supporting $\mathscr{S}$. ■

The inverse construction is not always possible. We already have seen that (4.32) entails that, for signalling games which fully support a set of support problems, $p(\theta|v)$ must be the same for all $v \in K_S$. Hence, there cannot be for every signalling game a set of support problems which is fully supported by it. If $\mathscr{G}$ is such that $p(\theta|v)$ is the same for all $v \in K_S^\theta := \{v \in \Omega \mid \mu(v|\theta) > 0\}$, then we can set $P_S^\theta(v) := P(v|K_S^\theta)$ and $P_H^\theta(v) := P(v|K_H^\theta)$ with $K_H^\theta := \{v \in \Omega \mid P(v) > 0\}$. Then K_H^θ and P_H^θ do not depend on θ, and we find $\mu(v|\theta) = P(v)\,p(\theta|v)/\sum_w (P(w)\,p(\theta|w)) = P(v)/P(K_S^\theta) = P(v|K_S^\theta) = P_H(v|K_S^\theta) = P_S^\theta(v)$.

For the general case, we either have to give up (4.32) or full reliability. If we decide to give up (4.32), then we can set $P_S^\theta = \mu(v|\theta)$ and e.g. $P_H(v) = P(v)$, and arrive for each θ at a support problem with fully reliable speaker expectations. If we decide to give up full reliability, then we can set $P_S^\theta(v) = P(v|K_S^\theta)$ and $P_H = P$, and arrive for each θ at a reliable support problem which satisfies (4.32). In either case, P_H does not depend on θ. Hence, the support problems in the constructed set $\mathscr{S}$ do only differ with respect to P_S.

We summarise the result:

Lemma 4.8 *Let $\mathscr{G} = \langle \Omega, \Theta, P, p, \mathscr{F}, \mathscr{A}, u \rangle$ be a given signalling game. For $\theta \in \Theta$, let $K_S^\theta := \{v \in \Omega \mid \mu(v|\theta) > 0\}$, $P_S^\theta(v) := P(v|K_S^\theta)$, and $P_H^\theta := P$. Let σ_θ be the resulting support problem. Then, the P_S^θ are reliable, and it holds:*

1. *the support problems σ_θ satisfy (4.32): $P_S^{\sigma_\theta} = P(v|K_S^\theta) = P_H(v|K_S^\theta)$.*
2. *If, in addition, $p(\theta|v)$ is the same for all $v \in K_S^\theta$, then the support problems σ_θ are also fully reliable, i.e. $P_S^{\sigma_\theta} = \mu(\,.\,|\theta)$.*

Support problems which do not satisfy (4.32) were considered in (Benz, 2006).

5 Implicatures

In this section, we apply the ideas of Section 3 to signalling games and prove more explicit characterisations of implicatures. We assume throughout that a fixed signalling game $\mathscr{G} = \langle \Omega, \Theta, P, p, \mathscr{F}, \mathscr{A}, u \rangle$ together with as strategy pair (S,H) is given. As explained in Section 3, $\langle \mathscr{G}, S, H \rangle$ defines a linear causal Bayesian network $(\mathscr{X}_i, p_i)_{i=0,\ldots,3}$ if we identify $\mathscr{X}_0$ with Ω, $\mathscr{X}_1$ with Θ, $\mathscr{X}_2$ with $\mathscr{F}$, and $\mathscr{X}_3$ with $\mathscr{A}$; accordingly, we set $p_0 = P$, $p_1 = p$, $p_2 = S$, and $p_3 = H$. In this section, we write $\mu(\theta)$ and $\mu(F)$ for the marginal probabilities $\mu_1(\theta)$ and $\mu_2(F)$, and $\mu(\theta|F)$ for the conditional marginal probability $\mu_{1|2}(\theta|F)$.[4]

[4]The definitions of these probability distributions in the form of explicit sums can be found in (2.3) and (6.52).

We write, by a small mis-use of logical notation, $\langle \mathscr{G}, S, H \rangle \models F +\!> R$ if the utterance of F implicates R. In Lemma 3.4, we have shown that for any set $Y \subseteq \mathscr{F}$ and $X \subseteq \mathscr{X}_i$ it holds that:

$$\langle \mathscr{G}, S, H \rangle \models Y +\!> X \iff \mu_{i|\mathscr{F}}(X|Y) = 1 \tag{5.38}$$

In traditional theories of implicatures, it is assumed that an implicature provides information about the world or the speaker's information state in addition to the literally communicated information. Therefore, we are now concentrating on the cases $\mathscr{X}_i = \Omega$ and $\mathscr{X}_i = \Theta$; i.e. we are looking for a characterisation of implicatures *about the world* and the *speaker's state*. For $F \subseteq \mathscr{F}$ with $\mu_{\mathscr{F}}(F) > 0$, and $R_0 \subseteq \Omega$ or $R_1 \subseteq \Theta$, the criterion in (5.38) reads as:

$$\langle \mathscr{G}, S, H \rangle \models F +\!> R_i \iff \mu(R_i|F) = 1. \tag{5.39}$$

By definition, $\mu(R_i|F) = 1$ is equivalent to

$$\frac{\mu(\pi_i^{-1}[R_i] \cap \pi_{\mathscr{F}}^{-1}[F])}{\mu(\pi_{\mathscr{F}}^{-1}[F])} = 1. \tag{5.40}$$

We first consider R_1, which is a subset of Θ. Then (5.40) is equivalent to $\{\theta \in R_1 \,|\, \mu_\Theta(\theta) > 0 \wedge S(F|\theta) > 0\} \supseteq \{\theta \in \Theta \,|\, \mu_\Theta(\theta) > 0 \wedge S(F|\theta) > 0\}$. If $\mu_\Theta(\theta) > 0$ for all $\theta \in \Theta$, then this formula is again equivalent to $\forall \theta : S(F|\theta) > 0 \Rightarrow \theta \in R_1$.

We now turn to the implicatures about the state of the world, i.e. to R_0, which is a subset of Ω. Then (5.40) is equivalent to $\{v \in R_0 \,|\, P(v) > 0 \wedge \exists \theta\, (p(\theta|v) > 0 \wedge S(F|\theta) > 0)\} \supseteq \{v \in \Omega \,|\, P(v) > 0 \wedge \exists \theta\, (p(\theta|v) > 0 \wedge S(F|\theta) > 0)\}$. If $P(v) > 0$ for all $v \in \Omega$, then this formula is again equivalent to $\forall v : \mu(F|v) > 0 \Rightarrow v \in R_0$.

We summarise this result in the following proposition:

Proposition 5.1 *Let $\mathscr{G} = \langle \Omega, \Theta, P, p, \mathscr{F}, \mathscr{A}, u \rangle$ be a signalling game and (S,H) a strategy pair. Let $F \subseteq \mathscr{F}$ with $\mu_{\mathscr{F}}(F) > 0$. Then it holds:*

1. *If $R \subseteq \Theta$, and if for all $\theta \in \Theta$ $\mu_\Theta(\theta) > 0$, then*

$$\langle \mathscr{G}, S, H \rangle \models F +\!> R \iff \forall \theta : S(F|\theta) > 0 \Rightarrow \theta \in R. \tag{5.41}$$

2. *If $R \subseteq \Omega$, and if for all $v \in \Omega$ $P(v) > 0$, then*

$$\langle \mathscr{G}, S, H \rangle \models F +\!> R \iff \forall v : \mu(F|v) > 0 \Rightarrow v \in R. \tag{5.42}$$

Note that the implicatures are completely independent of the meaning of the signals in $\mathscr{F}$. Hence, they are also defined for situations in which the signals have no pre-defined semantic meaning. The implicature of a signal coincides with Lewis notion of *indicated meaning* (2002). Lewis *defined* the semantic

meaning of signals as their indicated meaning. In this way, he could explain how the semantics of signals can emerge from a convention about their use. If we assume that a semantics is already established, then the indicated meaning may exceed this pre-defined semantic meaning. This additional information is commonly called an implicature. Our definition in (5.39) differs from common usage of the word *implicature* in so far as the literal meaning of a signal, if defined, is subsumed by implicated meaning. We can define a stronger notion of implicature which is more in accordance with the common usage. According to this notion, an utterance of F implicates R only if R does not already follow from the communicated semantic meaning of F. We only introduce this notion in order to show that an equivalent to the common notion of implicature can easily be derived from our definition; but the concept of proprt implicatures will not be used anywhere in this paper.

Definition 5.2 (Proper Implicatures) *Let $\mathscr{S}$ be a set of interpreted support problems $\langle \Omega, P_S, P_H, \mathscr{F}, \mathscr{A}, u, c, [\![\,.\,]\!] \rangle$ which may only differ with respect to P_S. Let (S,H) be a strategy pair for $\mathscr{S}$. For $R \subseteq \Omega$, $F \in \mathscr{F}$, and $[\![F]\!]^* := \{v \in [\![F]\!] \mid P_H(v) > 0\}$, we say that the utterance of F* properly implicates *that R in $\langle \mathscr{S}, S, H \rangle$ iff $\langle \mathscr{S}, S, H \rangle \models F +\!> R$ & $[\![F]\!]^* \setminus R \neq \emptyset$.*

We now turn our attention to support problems. In (Benz, 2008), the implicatures $R \subseteq \Omega$ of a sentence F in a given set of support problems $\mathscr{S}$ were defined by $\langle \mathscr{S}, S, H \rangle \models F +\!> R \iff \forall \sigma \in \mathscr{S}\,(F \in \mathrm{Op}_\sigma \Rightarrow P_S^\sigma(R) = 1)$. We now show that:

Lemma 5.3 *Let $\mathscr{S}$ be a set of support problems $\sigma = \langle \Omega, P_S, P_H, \mathscr{F}, \mathscr{A}, u, c, [\![\,.\,]\!] \rangle$ which only differ with respect to P_S^σ. Let $\mathscr{G}_0$ and $\mathscr{G}_1$ both be signalling games which support $\mathscr{S}$. Let (S,H) be a pair of signalling strategies for $\mathscr{G}_0$ and $\mathscr{G}_1$. Then, it holds:*

1. *$\mu^{\langle \mathscr{G}_0, S, H \rangle}(F) > 0$ iff $\mu^{\langle \mathscr{G}_1, S, H \rangle}(F) > 0$.*
2. *If $\mu^{\langle \mathscr{G}_i, S, H \rangle}(F) > 0$ and $R \subseteq \Omega$, then it holds:*

$$\langle \mathscr{G}_i, S, H \rangle \models F +\!> R \iff \forall \sigma \in \mathscr{S}\,(S(F|\sigma) > 0 \Rightarrow P_S^\sigma(R) = 1). \quad (5.43)$$

3. *If $\mu^{\langle \mathscr{G}_i, S, H \rangle}(F) > 0$ and $R \subseteq \Omega$, then it holds:*

$$\langle \mathscr{G}_0, S, H \rangle \models F +\!> R \iff \langle \mathscr{G}_1, S, H \rangle \models F +\!> R. \quad (5.44)$$

Proof: That the $\mathscr{G}_i$ support $\mathscr{S}$ implies, by Def. 4.4, that for all $v \in \Omega$: $P_S^\sigma(v) > 0 \Leftrightarrow P^{\mathscr{G}_i}(v)\, p^{\mathscr{G}_i}(\sigma|v) > 0$. By definition of $\mu(F)$ in (2.3), $\mu^{\langle \mathscr{G}_i, S, H \rangle}(F) > 0$ iff $\sum_{v,\sigma} P^{\mathscr{G}_i}(v)\, p^{\mathscr{G}_i}(\sigma|v)\, S(F|\sigma) > 0$. As the $\mathscr{G}_i$ support $\mathscr{S}$, the latter is equivalent to $\sum_{v,\sigma} P_S^\sigma(v)\, S(F|\sigma) > 0$. From this, the first claim follows immediately.

Let us now only consider $\mathscr{G}_0$. Let $\mu(v,\sigma,F) := P(v)\,p(\sigma|v)\,S(F|\sigma)$. Then, with (5.42) we find

$$\begin{aligned}\langle \mathscr{G}_0, S, H\rangle \models F +\!> R &\Leftrightarrow \forall v:\ \mu(F|v) > 0 \Rightarrow v \in R \qquad (5.45)\\ &\Leftrightarrow \{\langle v,\sigma,F\rangle \mid \mu(v,\sigma,F) > 0\} \subseteq \{\langle v,\sigma,F\rangle \mid \mu(v,\sigma,F) > 0 \wedge v \in R\}.\end{aligned}$$

As all P_S^σ are reliable, the last condition is equivalent to $\{\langle v,\sigma,F\rangle \mid P_S^\sigma(v) > 0 \wedge S(F|\sigma) > 0\} \subseteq \{\langle v,\sigma,F\rangle \mid P_S^\sigma(v) > 0 \wedge S(F|\sigma) > 0 \wedge v \in R\}$. This is again equivalent to $\forall v,\sigma\,(P_S^\sigma(v) > 0 \wedge S(F|\sigma) > 0 \Rightarrow v \in R)$, which finally is equivalent to $\forall\sigma\,(S(F|\sigma) > 0 \Rightarrow P_S^\sigma(R) = 1)$. This proves the second claim. The third claim follows immediately from the second. ■

With (5.44), we can define:

Definition 5.4 *Let $\mathscr{S}$ be a set of support problems σ which only differ with respect to P_S^σ. Let* Supp$(\mathscr{S})$ *be the set of all signalling games $\mathscr{G}$ which support $\mathscr{S}$. Let (S,H) be any strategy pair for $\mathscr{S}$, and let $F \in \mathscr{F}$ be such that $\exists\sigma \in \mathscr{S}\, S(F|\sigma) > 0$. Then we set for $R \subseteq \Omega$:*

$$\langle \mathscr{S}, S, H\rangle \models F +\!> R \iff \forall \mathscr{G} \in \text{Supp}(\mathscr{S})\ \langle \mathscr{G}, S, H\rangle \models F +\!> R. \qquad (5.46)$$

Note that by Lemma 4.7 Supp$(\mathscr{S})$ is never empty. If (S,H) is the *canonical solution* to $\mathscr{S}$, we arrive with (5.43) at:

$$\langle \mathscr{S}, S, H\rangle \models F +\!> R \iff \forall\sigma \in \mathscr{S}\,(F \in \text{Op}_\sigma \Rightarrow P_S^\sigma(R) = 1). \qquad (5.47)$$

Starting from (5.47), we can derive criteria for special but frequent situations. The remainder of the section presents some results from (Benz, 2008).

First, we note that, as the hearer has to check all support problems in $\mathscr{S}$, we arrive at the more implicatures the smaller $\mathscr{S}$ becomes. If $\mathscr{S} = \{\sigma\}$ and $F \in \text{Op}_\sigma$, then F will implicate everything the speaker knows. The other extreme is the case in which answers implicate only what they logically entail. We show in Proposition 5.7 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:

$$O(a) := \{w \in \Omega \mid \forall b \in \mathscr{A}\ u(w,a) \geq u(w,b)\}. \qquad (5.48)$$

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 $\mathscr{S}$:

Definition 5.5 (Expert) *Let $\mathscr{S}$ be a set of support problems with joint decision problem $\langle(\Omega,P_H),\mathscr{A},u\rangle$. Then we call S an* expert *in a support problem σ if $\exists a\in\mathscr{A}\,P_S^\sigma(O(a))=1$. He is an expert in $\mathscr{S}$, if he is an expert in every $\sigma\in\mathscr{S}$.*

This leads us to the following criterion for implicatures:

Lemma 5.6 *Let $\mathscr{S}$ be a set of support problems with joint decision problem $\langle(\Omega,P_H),\mathscr{A},u\rangle$, and (S,H) its canonical solution. Assume furthermore that E is an expert for every $\sigma\in\mathscr{S}$ and that $\forall v\in\Omega\,\exists\sigma\in\mathscr{S}\,P_S^\sigma(v)=1$. Let $\sigma\in\mathscr{S}$ and $F,R\subseteq\Omega$ be two propositions with $F\in\mathrm{Op}_\sigma$. Then, with $F^*:=\{v\in\Omega\,|\,P_H(v)>0\}$, it holds that:*

$$\langle\mathscr{S},S,H\rangle\models F+\!\!>R \text{ iff } F^*\cap\bigcap_{a\in\mathscr{B}(F)}O(a)\subseteq R. \tag{5.49}$$

Proof: We first show that

$$(\exists a\in\mathscr{A}\,P_S^\sigma(O(a))=1\,\&\,F\in\mathrm{Op}_\sigma)\Rightarrow\forall a\in\mathscr{B}(A):\,P_S^\sigma(O(a))=1. \tag{5.50}$$

Let a,b be such that $P_S^\sigma(O(a))=1$ and $P_S^\sigma(O(b))<1$. Then

$$\begin{aligned}EU_E^\sigma(b)=\sum_{v\in O(a)}P_S^\sigma(v)\cdot u(v,b)<&\sum_{v\in O(a)\cap O(b)}P_S^\sigma(v)\cdot u(v,a)\\&+\sum_{v\in O(a)\setminus O(b)}P_S^\sigma(v)\cdot u(v,a)=EU_E^\sigma(a).\end{aligned}$$

With $K_S=\{v\in\Omega\,|\,P_S^\sigma(v)>0\}$ it follows that $b\notin\mathscr{B}(K_S)$, and by (4.37) that $b\notin\mathscr{B}(A)$. Hence, $b\in\mathscr{B}(A)$ implies $P_S^\sigma(O(b))=1$.

Let $F^+:=\bigcap_{a\in\mathscr{B}(A)}O(a)$. We first show that $F^*\cap F^+\subseteq R$ implies $F+\!\!>R$. Let $\hat\sigma\in\mathscr{S}$ be such that $F\in\mathrm{Op}_{\hat\sigma}$. We have to show that $P_S^{\hat\sigma}(R)=1$. By (5.50) $P_S^{\hat\sigma}(F^+)=P_S^{\hat\sigma}(\bigcap_{a\in\mathscr{B}(A)}O(a))=1$ and by (4.32) $P_S^{\hat\sigma}(F^*)=1$; hence $P_S^{\hat\sigma}(F^+\cap F^*)=1$, and it follows that $P_S^{\hat\sigma}(R)=1$.

Next, we show $F+\!\!>R$ implies $F^*\cap F^+\subseteq R$. Suppose that $F^*\cap F^+\not\subseteq R$. Let $w\in F^*\cap F^+\setminus R$. From condition $\forall v\in\Omega\,\exists\hat\sigma\in\mathscr{S}\,P_S^{\hat\sigma}(v)=1$ it follows that there is a support problem $\hat\sigma$ such that $P_S^{\hat\sigma}(w)=1$. As $w\in F^+$, it follows by (4.37) that $F\in\mathrm{Op}_{\hat\sigma}$. Due to $F+\!\!>R$, it follows that $P_S^{\hat\sigma}(R)=1$, in contradiction to $w\notin R$. ■

F^* is the equivalent to the common ground updated with F. In the context of a support problem, we can interpret an answer F as a *recommendation* to choose one of the action in $\mathscr{B}(F)$. We may say that the recommendation is *felicitous* only if all recommended actions are optimal. Hence, F^+ 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 $\mathscr{B}(F)=\mathscr{B}(F^*)$ by Definition 4.30; hence $\bigcap_{a\in\mathscr{B}(F)}O(a)=\bigcap_{a\in\mathscr{B}(F^*)}O(a)$

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 $\mathscr{S}$ in which the conditions of Lemma 5.6 hold but in which answers only implicate what they entail:

Proposition 5.7 *Let $\mathscr{S}$ be a set of support problems with joint decision problem $\langle(\Omega, P_H), \mathscr{A}, u\rangle$. Let (S,H) be the canonical solution. Assume that for all $X \subseteq \Omega$, $X \neq \emptyset$: $\exists\sigma \in \mathscr{S}\, K_S^\sigma = X$ and $\exists a \in \mathscr{A}\, O(a) = X$. Then, for all $\sigma \in \mathscr{S}$ with $F \in \mathrm{Op}_\sigma$ it holds $\forall R \subseteq \Omega : \langle \mathscr{S}, S, H\rangle \models F +> R \Leftrightarrow F^* \subseteq R$.*

Proof: Condition $\forall X \neq \emptyset\, \exists a \in \mathscr{A}\, O(a) = X$ trivially entails that E is an expert for all $\sigma \in \mathscr{S}$. Condition $\forall X \neq \emptyset\, \exists\sigma \in \mathscr{S}\, K_S^\sigma = X$ entails the second condition of Lem. 5.6: $\forall v \in \Omega\, \exists\sigma \in \mathscr{S}\, P_S^\sigma(v) = 1$. Then, let $F \in \mathrm{Op}_\sigma$ and let a^* be such that $O(a^*) = F^*$; as $\mathscr{B}(F) = \mathscr{B}(F^*)$, it follows that $\bigcap\{O(a) \mid a \in \mathscr{B}(F)\} = \bigcap\{O(a) \mid a \in \mathscr{B}(F^*)\} = O(a^*) = F^*$. Hence, by Lem. 5.6, $F +> R$ iff $F^* \subseteq R$. ■

This proposition also shows that the conditions of Lemma 5.6 are less restrictive than they might seem to be.

6 The Fundamental Lemma

In this section we prove Lemma 2.3. For convenience, let us again repeat the definition of signalling games from Definition 2.1. A signalling game is a structure $\langle \Omega, \Theta, P, p, \mathscr{F}, \mathscr{A}, u\rangle$ for which: (1) Ω and Θ are non–empty finite sets; (2) $P(\,.\,)$ is a probability distribution over Ω; (3) $p(\,.\,|v)$ is a probability distribution over Θ for every $v \in \Omega$; (4) $\mathscr{F}$ and $\mathscr{A}$ are respectively the speaker's and hearer's action sets; and (5) $u : \Omega \times \Theta \times \mathscr{F} \times \mathscr{A} \to \mathbb{R}$ is a shared utility function which can be decomposed such that $u(v,\theta,F,a) = u(v,a) - c(F)$ for some strictly positive function $c : \mathscr{F} \to \mathbb{R}^+$.

We first introduce Bayesian perfect equilibria and then prove Lemma 2.3. As mentioned before, it can be more convenient to calculate the Bayesian perfect equilibria than the Nash equilibria of a signalling game.

Definition 6.1 (Perfect Bayesian Equilibrium) *A strategy pair (S,H) is a* perfect Bayeasian equilibrium *of a signalling game $\langle \Omega, \Theta, P, p, \mathscr{F}, \mathscr{A}, u\rangle$ iff:*

1. *For all S' and all θ with $\mu(\theta) > 0$ it is $\mathscr{E}_S(S'|\theta) \leq \mathscr{E}_S(S|\theta)$,*
2. *For all H' and all F with $\mu(F) > 0$ it is $\mathscr{E}_H(H'|F) \leq \mathscr{E}_H(H|F)$.*

The equilibrium is strict *if we can replace $\leq$ by $<$. It is* weak *if it is not strict.*

We show that the Bayesian perfect equilibria are the same as Nash equilibria in the sense of Definition 2.2. For this we show that a hearer strategy H is a best response to a speaker strategy S iff $\mathscr{E}_H(H|F)$ is maximal for each F with

non–negative probability. For the following calculations, it should be noted that the payoff function u does not depend on θ. Hence, we could arbitrarily choose a θ_0 and keep it as a fixed argument of u. With $\mu(F)$ defined as before, it is:

$$\begin{aligned}
\mathscr{E}(H|S) &= \sum_{v\in\Omega} P(v) \sum_{\theta\in\Theta} p(\theta|v) \sum_{F\in\mathscr{F}} S(F|\theta) \sum_{a\in\mathscr{A}} H(a|F)\, u(v,\theta,F,a) \\
&= \sum_F \mu(F) \sum_v \sum_a H(a|F)\, u(v,\theta_0,F,a)\, \frac{P(v)\sum_{\theta\in\Theta} p(\theta|v)\, S(F|\theta)}{\mu(F)} \\
&= \sum_F \mu(F) \sum_v \mu_H(v|F) \sum_a H(a|F)\, u(v,\theta,F,a) \\
&= \sum_{F\in\mathscr{F}} \mu(F)\, \mathscr{E}_H(H|F). \qquad (6.51)
\end{aligned}$$

Hence, for fixed speaker strategy S, $\mathscr{E}(H|S)$ becomes maximal iff $\mathscr{E}_H(H|F)$ is maximal for all F with $\mu(F) > 0$, i.e. for all F for which the probability of being received by the hearer is greater zero. Similarly, it can be shown that $\mathscr{E}(S|H) = \sum_\theta \mu(\theta)\, \mathscr{E}_S(S|\theta)$. Hence, the Bayesian perfect equilibria in the sense of Definition 6.1 are identical to the Nash equilibria in the sense of Definition 2.2.

We now turn to the proof of Lemma 2.3. For this, we first reformulate the hearer's expected utility in terms of the conditional probability of the speaker's type being θ given answer F. This allows us to derive an estimate of the maximal expected utility. Hence, let us consider the expected utility $\mathscr{E}_H(H|F)$ of a hearer strategy H after receiving signal F. With (2.2) and (2.4), we find:

$$\begin{aligned}
\mathscr{E}_H(H|F) &= \sum_a H(a|F)\, \frac{\sum_v P(v)\, \sum_\theta p(\theta|v)\, S(F|\theta)}{\mu(F)}\, u(v,\theta,F,a) \\
&= \frac{1}{\mu(F)} \sum_\theta S(F|\theta) \sum_a H(a|F) \sum_v P(v)\, p(\theta|v)\, u(v,\theta,F,a) \\
&= \frac{1}{\mu(F)} \sum_\theta S(F|\theta)\, \mu(\theta) \sum_a H(a|F)(\mathscr{E}_S(a|\theta) - c(F)) \\
&= \sum_\theta \frac{S(F|\theta)\, \mu(\theta)}{\mu(F)} \sum_a H(a|F)\, \mathscr{E}_S(a|\theta) - c(F)
\end{aligned}$$

Let's write

$$\mu_{\Theta|\mathscr{F}}(\theta|F) = \frac{S(F|\theta)\, \mu(\theta)}{\mu(F)} = \frac{S(F|\theta) \sum_w P(w)\, p(\theta|w)}{\sum_\theta S(F|\theta) \sum_w P(w)\, p(\theta|w)}. \qquad (6.52)$$

This is the hearer's probability of the speaker type being θ given F. In the following, we will also use the short form $\mu(\theta|F)$ for $\mu_{\Theta|\mathscr{F}}(\theta|F)$. With this

abbreviation, we can summarise the result as follows:

$$\mathscr{E}_H(H|F) = \sum_{\theta} \mu(\theta|F) \sum_{a} H(a|F)\, \mathscr{E}_S(a|\theta) - c(F) \tag{6.53}$$

Let $M_\theta := \max_a \mathscr{E}_S(a|\theta)$. This is the maximal expected utility given θ. An action is *optimal* given θ if its expected utility is maximal. Hence, $\mathscr{E}_S(a|\theta) = M_\theta$ iff a is an element of the set $\mathscr{B}(\theta)$ of all actions with maximal expected utility, which has been defined in (2.8) as:

$$\mathscr{B}(\theta) = \{a \in \mathscr{A} \mid \forall b \in \mathscr{A}\, \mathscr{E}_S(b|\theta) \leq \mathscr{E}_S(a|\theta)\}. \tag{6.54}$$

Hence, for fixed θ we find:

1. If $H(a|F) > 0 \Rightarrow a \in \mathscr{B}(\theta)$, then

$$\sum_{a} H(a|F)\, \mathscr{E}_S(a|\theta) = M_\theta. \tag{6.55}$$

2. If $\exists a \notin \mathscr{B}(\theta)\, H(a|F) > 0$, then

$$\sum_{a} H(a|F)\, \mathscr{E}_S(a|\theta) < M_\theta. \tag{6.56}$$

It follows then from (6.53) that a strategy H is guaranteed to be optimal if $\mu(\theta|F) > 0$ entails for all θ that $(H(a|F) > 0 \Rightarrow a \in \mathscr{B}(\theta)))$, i.e. $H(\mathscr{B}(\theta)|\theta) = 1$. As mentioned before, we implicitly assume that in (6.52) the denominator $\mu(F)$ of μ is greater zero. Hence, if $\mu(\theta) > 0$, i.e. if θ is assigned to the speaker with a positive probability, then it follows that $\forall\theta(S(F|\theta) > 0 \Rightarrow H(\mathscr{B}(\theta)|F) = 1)$ entails that $\mathscr{E}_H(H|F)$ is maximal.

These considerations lead to the following criteria. Let Θ^* be the set of all types θ for which $\exists v\, P(v)\, p(\theta|v) > 0$, and let $F \in \mathscr{F}$. Let (S,H) be a strategy pair which satisfies the following condition:

$$\forall\theta \in \Theta^*\, (S(F|\theta) > 0 \Rightarrow H(\mathscr{B}(\theta)|F) = 1). \tag{6.57}$$

Then it follows that H is a best response to F, i.e. for all hearer strategies H' it holds that $\mathscr{E}_H(H'|F) \leq \mathscr{E}_H(H|F)$. Furthermore, if H' is such that

$$\exists\theta \in \Theta^*\, \exists a \notin \mathscr{B}(\theta)\, (S(F|\theta) > 0 \wedge H'(a|F) > 0), \tag{6.58}$$

then $\mathscr{E}_H(H'|F) < \mathscr{E}_H(H|F)$. Equations (6.57) and (6.58) entail Lemma 2.3.

7 Implicatures and Ambiguity

In this section, we address a problem that is shared by all accounts that assume that disambiguation is achieved by maximising expected utilities of interpretations. This method of disambiguation is the central principle for explaining pragmatic phenomena in Prashant Parikh's framework of games of partial information (2001). His standard example is the following sentence showing a scope ambiguity:

(2) **a)** Every ten minutes a man gets mugged in New York. (A)

b) Every ten minutes some man or other gets mugged in New York. (F)

c) Every ten minutes a particular man gets mugged in New York. (F')

The sentence A is ambiguous between the interpretation in which it is always the same person which gets mugged (R'), and the interpretation in which it is a random sequence of people who gets mugged (R). Speaker and hearer have to coordinate their strategies such that the hearer arrives at the interpretation that the speaker had in mind. With F being the unambiguous sentence with meaning R, F' the unambiguous sentence with meaning R', and ρ, ρ' the probabilities of R, R' respectively, we arrive at the game tree shown in Figure 4. First nature

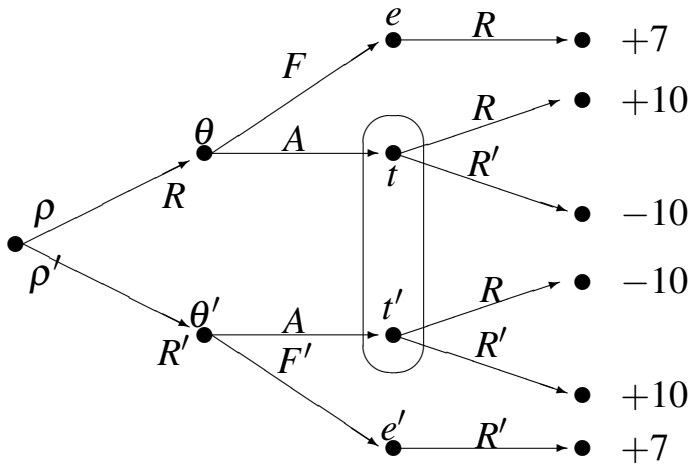


Fig. 4: Parikh's game tree for Example (2).

chooses between R and R'. If it has chosen R, then the speaker in situation θ has the choice between the unambiguous but more complex F and the ambiguous but simpler A. If nature has chosen R', then the speaker in situation θ' has the choice between the unambiguous but more complex F' and again the ambiguous but simpler A. If the speaker chooses F or F', then there is only one

interpretation which the hearer can choose. If the speaker choses A, then the ambiguity of A leads to the choice between two different interpretations. The numbers at the end of the branches denote the shared utilities of speaker and hearer.

Prashant Parikh solves this game by calculating the Nash equilibria, and, if there is more than one Nash equilibrium, choosing the equilibrium which leads to the higher overall expected payoff, the so–called *Pareto* Nash equlibrium. It is easy to see that there are exactly two Nash equilibria in the situation of Figure 4: One Nash equilibrium (S,H) in which the speaker chooses F in θ and A in θ', and in which the hearer interprets A by R'; another Nash equilibrium (S',H') in which the speaker chooses F' in θ' and A in θ, and in which the hearer interprets A by R. As the probability ρ of it being always the same man who gets mugged is much lower than the probability ρ', the first strategy will more often avoid the use of the complex formula F', and hence lead to a higher overall expected utility. Hence, the first strategy pair (S,H) is the unique Pareto Nash equilibrium of this game, and the hearer will interpret A as meaning R. According to Parikh, this shows that the utterance of A *communicates with certainty* that R (Parikh, 1990), (Parikh, 2006)[p. 104].

Implicatures are explained by Parikh (2001) along the same lines. He assumes that an utterance is ambiguous between the literal meaning (A) and the literal meaning + implicature $(A+R)$. The implicature $A+\!> R'$ is explained by the fact that for the Pareto Nash equilibrium (S,H) which solves the resulting game it holds that $H(A) = R'$. This account is principally different from the account provided in the Optimal Answer model. There, the solution (S,H) is calculated by backward induction,[5] and the implicature is identified with the additional information that an utterance A provides about the speaker's information state, i.e. with $S^{-1}(A)$. But, although the two approaches differ here, the same predictions about disambiguation are made in the Optimal Answer model and in Parikh's model.

Our principal counterexample against the idea that ambiguities are resolved by choosing the more probable interpretation, and hence that this interpretation is thereby communicated with *certainty*, is the Doctor's Appointment example:

(3) John is known to regularly consult two different doctors, physicians A and B. He consults A more often than B. S meets H and tells him:

S: *John has a doctor's appointment at 4pm. He requests you to pick him up afterwards.* (D)

[5] As we have shown in Lemma 4.5, the solution found by backward induction is always a Pareto Nash equilibrium.

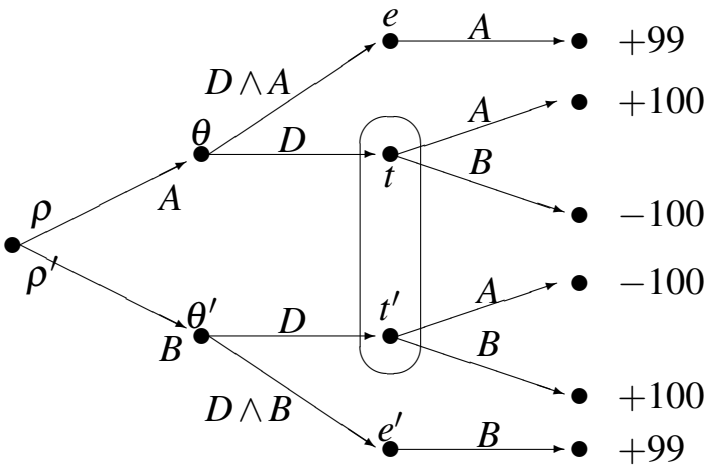


Fig. 5: The game tree for the Doctor's Appointment example.

Clearly, *S* fails to communicate that John waits at A's practice. Structurally, the situation is identical to the situation shown in Figur (4). In Figure 5, we see the game tree of the Doctor's Appointment example. The hearer has the choice between two different interpretations: *A* that John is waiting at A's practice, and *B* that John is waiting at B's practice. Hence, if rational interlocutors resolve the ambiguity by choosing the more probable interpretation in Figur (4), then in (3) they should resolve it in the same way. But in this case, *D* would have to communicate *with certainty* that John has an appointment with physician A, which is clearly not the case.

The problem does not only arise with Parikh's model. In the Doctor's Appointment example, backward induction predicts that *D* is an optimal assertion if the speaker knows that John is at A's practice but not if he knows that John is at B's practice. Hence, we are faced with the problem to explain why *D* is not an optimal assertion in the Doctor's Appointment example. This means, we have to explain why backward induction is ruled out as a principle of disambiguation in the Doctor's Appointment scenario. This problem leads us to the consideration of games with *noisy communication* and *efficient clarification requests*.

Let us consider the addressee's natural reaction in the Doctor's Appointment example (3). What would be the natural response to the directive (D) *John has a doctor's appointment at 4pm. He requests you to pick him up afterwards*? Most probably, the addressee would just ask where John is waiting, at A's or at B's practice. If the answering person *S* is cooperative and knows about John's whereabouts, then he will tell *H* where to pick up John. Hence, the natural response to the ambiguity is a clarification request **c** which will induce *S* to provide an answer which allows *H* to unambiguously choose an optimal action *a* afterwards. We call such clarification requests *efficient* if they come with low costs and force the speaker to provide an unambiguously optimal answer. We

add efficient clarification requests in the next section to our models.

8 Efficient clarification requests

Until now, our signalling games modelled situations in which the hearer has immediately to decide which action to choose after receiving an answer from the speaker. In this section, we add efficient clarification requests to the hearer's action set. In the context of noisy communication, the possibility to ask clarification requests has a considerable effect on the equilibria of a game.

Assume that the hearer follows a strategy H which does not involve any clarification requests. Assume further that the hearer knows the speaker strategy S and receives an answer A. We now consider the question: When is it reasonable for the hearer to change his strategy $H(\,.\,|A)$ and ask a clarification request? For answering this question, we consider the set $Z(A)$ which consists of all pairs $\langle v, \theta \rangle$ which have positive probability, for which the speaker might answer A, and for which this might lead the hearer to choose a sub–optimal action. Clearly, for being optimal, S and H should be such that $Z(A)$ is empty. If it is not empty, the hearer can make it empty by changing H in such a way that he asks a clarification request whenever answer A occurs. We can consider this to be a *local* change in the sense that it only changes the hearer's strategy for answer A but leaves it unchanged for all other answers. Before we define local changes, we first introduce signalling games with *efficient clarification requests*:

Definition 8.1 *Let $\mathscr{G} = \langle \Omega, \Theta, P, p, \mathscr{F}, \mathscr{A}, u \rangle$ be a basic signalling game in the sense of Def. 2.1. Then $\mathscr{G}$ is a* signalling game with efficient clarification requests *iff there exists an act $\mathbf{c} \in \mathscr{A}$ and a cost function $c : \mathscr{F} \cup \{\mathbf{c}\} \to \mathbb{R}^+$ which satisfy the following conditions:*

1. (Efficiency): $u(v, \theta, F, \mathbf{c}) = \mu(\theta)\, M_\theta - (c(F) + c(\mathbf{c}))$, *with $M_\theta := \max_{a \in \mathscr{A} \setminus \{\mathbf{c}\}} \mathscr{E}_S(a|\theta)$ and $\mu(\theta)$ as in (2.3);*
2. (Nominality): *the cost function c is nominal;*
3. (Avoid $\mathbf{c}$): $\forall A, B \in \mathscr{F} : |c(A) - c(B)| < c(\mathbf{c})$.

M_θ is the maximal expected utility given θ. Hence, *efficiency* means that $\mathbf{c}$ achieves the maximal expected utility minus the costs of asking a clarification request. *Nominality* entails that these costs are positive but arbitrarily small in comparison to the utility of other actions. We will provide an exact definition of *nominality* in Section 11. (Avoid $\mathbf{c}$) says that if the speaker can find a more complex answer B which avoids a clarification request, then he will choose B rather than sticking to an answer A which he would have chosen otherwise.

As $\mathscr{B}(\theta) = \{a \in \mathscr{A} \mid \mathscr{E}_S(a|\theta) = M_\theta\}$, it follows that $\mathbf{c} \notin \mathscr{B}(\theta)$. With μ as in (6.52), it also follows that:

$$\mathscr{E}_H(\mathbf{c}|A) = \sum_\theta \mu(\theta|A)\, M_\theta - (c(A) + c(\mathbf{c})). \tag{8.59}$$

The proof is completely parallel to that of (6.53). As signalling games with efficient clarification requests are special cases of basic signalling games, it also follows that Lemma 2.3 remains valid.

Now we consider a situation in which speaker and hearer follow some strategy pair (S,H) which may violate condition (6.57). How can the hearer modify his strategy in order to achieve the best result? Assume that he receives answer A, then if there is a possibility that strategy H chooses a sub–optimal action, it is better for the hearer to ask a clarification request. This is the case if the following set $Z(A)$ is not empty:

$$Z(A) := \{\langle v, \theta\rangle \mid P(v)\, p(\theta|v)\, S(A|\theta) > 0 \wedge \exists a \notin \mathscr{B}(\theta)\, H(a|A) > 0\}. \tag{8.60}$$

$Z(A)$ is the set of all pairs of worlds v and speaker's types θ which have non–zero probability, for which the speaker answers A with non–zero probability, and for which H may choose a suboptimal action. It is convenient to collect the worlds v and the types θ that belong to $Z(A)$ in two separate sets:

$$Z^1(A) := \{v \in \Omega \mid \exists \theta\ \langle v, \theta\rangle \in Z(A)\}, \text{ and } Z^2(A) := \{\theta \in \Theta \mid \exists v\ \langle v, \theta\rangle \in Z(A)\}.$$

We show that if $Z(A) \neq \emptyset$, then the hearer strategy H is strictly dominated by a strategy $H^A_{\mathbf{c}}$ which is defined as follows:

$$H^A_{\mathbf{c}}(a|B) := \begin{cases} H(a|B) & \text{for } B \neq A \\ 1 & \text{for } a = \mathbf{c} \text{ and } B = A \end{cases}. \tag{8.61}$$

The strategy $H^A_{\mathbf{c}}$ is identical to H for all answers except A. For A it chooses a clarification request. We show the following proposition:

Proposition 8.2 *Let (S,H) be any strategy pair of a given signalling game $\mathscr{G}$ with efficient clarification request $\mathbf{c}$. Assume that $\mu_{\mathscr{F}}(A) > 0$, see (2.3). With $\mu(v,\theta) := P(v)\, p(\theta|v)$, μ_H as in (2.2), and $\mu_{\Theta|\mathscr{F}}$ as in (6.52), the following equivalences hold:*

$$\begin{aligned} Z(A) \neq \emptyset &\Leftrightarrow \mu(Z(A)) > 0 \Leftrightarrow P(Z^1(A)) > 0 \Leftrightarrow \mu_H(Z^1(A)|A) > 0 \\ &\Leftrightarrow \mu_{\Theta|\mathscr{F}}(Z^2(A)|A) > 0. \end{aligned} \tag{8.62}$$

It holds:

1. *If $Z(A) \neq \emptyset$, then $H^A_{\mathbf{c}}$ strictly dominates H.*
2. *If $Z(A) = \emptyset$, then H strictly dominates $H^A_{\mathbf{c}}$.*

Proof: The equivalences follow by unfolding the definitions. By (6.53) it holds that $\mathscr{E}_H(H|A) = \sum_\theta \mu_{\Theta|\mathscr{F}}(\theta|A) \sum_a H(a|A)\, \mathscr{E}_S(a|\theta) - c(A)$. We can split the sum $\sum_\theta$ into $\sum_{\theta \in Z^2(A)}$ and $\sum_{\theta \notin Z^2(A)}$. Hence, if $Z(A) = \emptyset$, and therefore $\mu_{\Theta|\mathscr{F}}(Z^2(A)|A) = 0$, then

$$\begin{aligned}\mathscr{E}_H(H|A) &= \sum_{\theta \notin Z^2(A)} \mu_{\Theta|\mathscr{F}}(\theta|A)\, M_\theta - c(A) \\ &> \sum_{\theta \notin Z^2(A)} \mu_{\Theta|\mathscr{F}}(\theta|A)\, M_\theta - (c(A) + c(\mathbf{c})) = \mathscr{E}_H(H^A_\mathbf{c}|A).\end{aligned}$$

Hence, H strictly dominates $H^A_\mathbf{c}$. Next, assume that $Z(A) \neq \emptyset$. This entails that:

$$\sum_\theta \mu_{\Theta|\mathscr{F}}(\theta|A) \sum_a H(a|A)\, \mathscr{E}_S(a|\theta) < \sum_\theta \mu_{\Theta|\mathscr{F}}(\theta|A)\, M_\theta$$

Hence, it follows with (Nominality) that

$$\begin{aligned}\mathscr{E}_H(H|A) &= \sum_\theta \mu_{\Theta|\mathscr{F}}(\theta|A) \sum_a H(a|A)\, \mathscr{E}_S(a|\theta) - c(A) \\ &< \sum_\theta \mu_{\Theta|\mathscr{F}}(\theta|A)\, M_\theta - (c(A) + c(\mathbf{c})) = \mathscr{E}_H(H^A_\mathbf{c}|A).\end{aligned}$$

Hence, $H^A_\mathbf{c}$ strictly dominates H. ■

The proposition shows how the hearer can improve his strategy for answer A without changing his strategy for answers different from A. We will exploit this property in the next section. In the same section, we will also see that a clarification request is not always the best response in situations in which the old strategy H would lead to sub–optimal choices. Proposition 8.2 only says that reacting with a clarification request is better than sticking to a faulty strategy, but there may be other possibilities to improve the old strategy. For achieving this result, we have to have a closer look at $Z(A)$. But this is more interesting in the context of noisy speaker strategies. We introduce models with expected noise in the next section, and will take up our consideration of $Z(A)$ in Section 10.

9 Expected Noise

There exist quite a number of equilibrium refinements in game theory which try to spell out which equilibria are stable under the assumption that strategies are noisy. Among the most widely discussed equilibria which deal with noisy strategies are *trembling hand perfect* equilibria (Selten, 1975) and *proper* equilibria (Myerson, 1978). 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. For proper equilibria it is assumed that the probability of mistakes depends on how good an action is. In our context, this means for a perturbed speaker strategy $\tilde{S}$ that the speaker chooses an answer which is just second to an optimal answer with probability at most ε times the probability of an optimal answer, and an answer that is third to an optimal answers with a probability ε times the probability of an second best answer, etc. For both criteria, the probability and the kind of mistakes can be inferred from theory *internal* parameters, as e.g. from the set of available hearer actions, their expected utilities, and the speaker's set of signals. In linguistic pragmatics, in contrast to other applications of game theory, the phenomena are very close to the cognitive level. Hence a strong interaction between the behavioural level, represented by game theory, and the cognitive level is to be expected. We introduce expected noise models as a framework to introduce noise into the game theoretic models which is controlled by *external* causes. The representation of noise in expected noise models is therefore very little restricted. In other respects, we simplify the model by only considering perturbations of the speaker's strategy. This means, we always assume that the hearer finds his best response with certainty.

In order to motivate the following definition, assume that an interpreted signalling game σ is given. Assume further that the speaker follows strategy S. Then $S(\,.\,|\sigma)$ will assign non–zero probability to certain forms $F \in \mathscr{F}$. They may form a proper sub–set $\mathscr{O}$ of $\mathscr{F}$. We may call S^ε a noisy ε-approximation of S if $\sum_F |S(F|\sigma) - S^\varepsilon(F|\sigma)| = \varepsilon$. Then, for $S^\varepsilon(\,.\,|\sigma)$ we can also collect all forms to which S^ε assign non–zero probability in a set $\mathscr{N}^\varepsilon$. The exact value of ε does not matter to us; hence, we abstract away from it and just keep the set of forms to which the S^ε assign non–zero probability. We assume that it is the same set for all ε. We call this set a *noise set*. Now, as we want to capture by these noise sets the perturbations resulting from cognitive sources, these sets may vary from support problem to support problem as the speaker's state varies from support problem to support problem. We therefore represent the perturbations by a function which maps $\mathscr{S}$ to sets $\mathscr{N}_\sigma \subseteq \mathscr{F}$. This motivates the following definition:

Definition 9.1 (EN model) *Let $\mathscr{S}$ be a set of interpreted support problems. Assume that the support problems $\langle \Omega, P_S, P_H, \mathscr{F}, \mathscr{A}, u, c, [\![\,.\,]\!]\rangle$ may only differ with respect to P_S. A model with* expected noise*, or* EN model*, is a triple $\langle \mathscr{S}, (\mathscr{O}_\sigma)_{\sigma\in\mathscr{S}}, (\mathscr{N}_\sigma)_{\sigma\in\mathscr{S}}\rangle$ for which*

1. *$(\mathscr{O}_\sigma)_{\sigma\in\mathscr{S}}$ is a sequence of sets $\mathscr{O}_\sigma \subseteq \mathscr{F}$.*

2. $(\mathscr{N}_\sigma)_{\sigma\in\mathscr{S}}$ *is a sequence of sets* $\mathscr{N}_\sigma \subseteq \mathscr{F}$.

In the following, we write $\langle \mathscr{S}, \mathscr{O}_\sigma, \mathscr{N}_\sigma \rangle$ instead of $\langle \mathscr{S}, (\mathscr{O}_\sigma)_{\sigma\in\mathscr{S}}, (\mathscr{N}_\sigma)_{\sigma\in\mathscr{S}} \rangle$. In our applications, $\mathscr{O}_\sigma$ is the set Op_σ of optimal answers of the canonical solution to the support problem σ.

If the hearer cannot distinguish between the elements of $\mathscr{S}$, then learning that the speaker produced a possibly noisy form F only provides him with the information that F is an element of the union of the $\mathscr{N}_\sigma$. Hence, we introduce the set:

$$\mathscr{N} := \bigcup_\sigma \mathscr{N}_\sigma. \tag{9.63}$$

It can be easily seen that the addition of efficient clarification requests in itself has no effect on the canonical solution. This changes when we consider *noisy communication*. We will see that the addition of noise and the availability of efficient clarification requests gives rise to a transformation $(\bar{S}, \bar{H})$ of the canonical solution (S,H) which is Pareto dominating all other strategies. In addition, the transformed solution is robust against the noise characterised by an EN model. A central role will be played by the sets $\tilde{\mathscr{B}}(A)$ and $\mathscr{F}_{\mathfrak{En}}$. $\tilde{\mathscr{B}}(A)$ is the set of all actions a which are optimal for all support problems σ for which A can occur as a noisy form:

$$\tilde{\mathscr{B}}(A) := \bigcap \{ \mathscr{B}(K_\sigma) \,|\, A \in \mathscr{N}_\sigma \} \text{ with } K_\sigma = \{ v \,|\, P_S^\sigma(v) > 0 \}. \tag{9.64}$$

$\mathscr{F}_{\mathfrak{En}}$ then collects all $A \in \mathscr{F}$ for which $\tilde{\mathscr{B}}(A)$ is not empty:

$$\mathscr{F}_{\mathfrak{En}} := \{ A \in \mathscr{N} \,|\, \tilde{\mathscr{B}}(A) \neq \emptyset \}. \tag{9.65}$$

We illustrate the meaning of these sets by a little example. Assume there are two support problems σ and σ'. Assume further that actions a and b are optimal in σ, and actions b and c are optimal in σ'. *Being optimal* has to be understood relative to the speaker's expectations P_S^σ; hence, we mean by saying that a and b are optimal in σ that $EU_S^\sigma(a) = EU_S^\sigma(b) = \max\{EU_S^\sigma(a') \,|\, a' \in \mathscr{A}\}$, and therefore *being optimal* is equivalent to being an element of $\mathscr{B}(K_\sigma)$. Hence, it is $\mathscr{B}(K_\sigma) = \{a,b\}$ and $\mathscr{B}(K_{\sigma'}) = \{b,c\}$. Let us now assume that $S^\varepsilon(A|\sigma) > 0$ and $S^\varepsilon(A|\sigma') > 0$. What is the best response for the hearer? If he chooses a, then this may be sub–optimal as he cannot be sure that the actual support problem is really σ. The same problem arises with c. But if he chooses b, then he is save as b is an optimal action in both σ and σ'. Choosing b is also better than asking a clarification request as this comes with additional (nominal) costs. Now, in contrast, consider a situation in which actions a and b are optimal in σ, and actions c and d are optimal in σ'. If again $S^\varepsilon(A|\sigma) > 0$ and $S^\varepsilon(A|\sigma') > 0$, then it is now better to ask a clarification request as there is no best choice which

belongs to all $\mathscr{B}(K_\sigma)$. Hence, we see that if $\mathscr{F}_{\mathfrak{En}} \neq \emptyset$, then the hearer can safely choose an act in $\mathscr{F}_{\mathfrak{En}}$, otherwise he should react with a clarification request.

As mentioned before, $\mathscr{O}_\sigma$ and $\mathscr{N}_\sigma$ are representing speaker strategies S and their perturbed forms S^ε. The following definition makes explicit in which sense EN models represent these strategies:

Definition 9.2 *Let $\mathfrak{En} = \langle \mathscr{S}, \mathscr{O}_\sigma, \mathscr{N}_\sigma \rangle$ be an EN model with $\mathscr{S}$ a set of support problems $\sigma = \langle \Omega, P_S, P_H, \mathscr{F}, \mathscr{A}, u, c, [\![.]\!] \rangle$. For $X \subseteq \mathscr{F}$, we denote by Δ^*_X the set of all completely mixed strategies over X, i.e. Δ^*_X is the set of all probability distributions P over X for which $P(x) > 0$ iff $x \in X$. Then, we say that:*

1. $\mathfrak{En}$ represents *a strategy S iff for all $\sigma \in \mathscr{S}$ $S(\,.\,|\sigma) \in \Delta^*_{\mathscr{O}_\sigma}$;*
2. $\mathfrak{En}$ represents *a noise strategy S^ε iff for all $\sigma \in \mathscr{S}$ $S^\varepsilon(\,.\,|\sigma) \in \Delta^*_{\mathscr{N}_\sigma}$;*
3. *an arbitrary strategy $\tilde{S}$ is an $\mathfrak{En}$ strategy iff for all $\sigma \in \mathscr{S}$*

$$\tilde{S}(\,.\,|\sigma) \in \Delta^{\mathfrak{En}}_\sigma := \Delta^*_{\mathscr{O}_\sigma} \cup \Delta^*_{\mathscr{N}_\sigma} \cup \Delta^*_{\mathscr{F}_{\mathfrak{En}}}. \tag{9.66}$$

Expected noise models are extensions of interpreted support problems. They represent the subjective level. Signalling games model the objective level, especially, objective success of communication is only definable for signalling games. Hence, in the following, we have again to consider the relation between signalling games and expected noise models. We repeat the definitions of the most important relations:

Definition 9.3 *A signalling game $\mathscr{G}$* supports *an EN model $\mathfrak{En} = \langle \mathscr{S}, \mathscr{O}_\sigma, \mathscr{N}_\sigma \rangle$ iff $\mathscr{G}$ supports $\mathscr{S}$. By Definition 4.4, this means that $\mathscr{G} = \langle \Omega, \Theta, P, p, \mathscr{F}, \mathscr{A}, u \rangle$ is such that $\Theta = \mathscr{S}$ and for all $\sigma = \langle \Omega, P_S, P_H, \mathscr{F}, \mathscr{A}, u, c, [\![.]\!] \rangle$ it is $\mu_\Theta(\sigma) = \sum_v P(v)\, p(\sigma|v) > 0$. We say that $\mathscr{G}$* fully *supports $\mathfrak{En}$ iff all P^σ_S are fully reliable; it* reliably *supports $\mathfrak{En}$ iff all P^σ_S are reliable; and it* weakly *supports $\mathfrak{En}$ iff all P^σ_S are truth preserving.*

9.1 The canonical solution

We consider situations in which the speaker may follow some perturbed strategy S^ε, and the hearer a strategy H. There may exist a support problem σ for which the speaker choose an answer A with probability greater zero to which the hearer may respond with a sub-optimal action if he follows H. Assume that this perturbed strategy S^ε is known to the hearer, and that he is in a situation in which he receives answer A. How does the hearer have to change his strategy $H(\,.\,|A)$ in order to achieve maximal expected payoff? We introduce an operation which changes the hearer's response to A but leaves it unchanged for all

other answers:

$$H_X^A(a|B) := \begin{cases} H(a|B) & \text{if } B \neq A, \\ |X|^{-1} & \text{if } B = A \wedge a \in X \end{cases}. \tag{9.67}$$

This operation turns strategy H into a strategy H_X^A which is identical to H for all answers except A, and for A it chooses each of the elements of X with equal probability. The strategy $H_{\mathbf{c}}^A$ defined in (8.61) is the special case in which the old response to A is replaced by the clarification request $\mathbf{c}$. We also consider the case in which the old response to A is replaced by $\tilde{\mathscr{B}}(A)$. We write:

$$H_{\mathbf{c}}^A := H_{\{\mathbf{c}\}}^A \text{ and } H^A := H_{\tilde{\mathscr{B}}(A)}^A. \tag{9.68}$$

It is quite intuitive that the hearer can optimise his strategy by changing $H(\,.\,|A)$ to $H_{\mathbf{c}}^A$ if $\tilde{\mathscr{B}}(A) = \emptyset$, and by changing it to H^A if $\tilde{\mathscr{B}}(A) \neq \emptyset$. If (S,H) is the canonical solution to $\mathscr{S}$, then, by applying these operations systematically, we arrive at a new *canonical solution* to the expected noise model. Its definition is provided in (9.69) and (9.71).

Definition 9.4 (Canonical Solution) *Let $\mathscr{S}$ be a set of support problems with canonical solution (S,H). Let $\mathfrak{En} = \langle \mathscr{S}, \mathscr{O}_\sigma, \mathscr{N}_\sigma \rangle$ be an expected noise model which represents S. Then, we define the* canonical extension $(\bar{S},\bar{H})$ *to $\mathfrak{En}$ as follows:*

$$\bar{H}(\,.\,|A) = \begin{cases} H^A, & \text{if } A \in \mathscr{F}_{\mathfrak{En}}, \\ H_{\mathbf{c}}^A & \text{if } A \notin \mathscr{F}_{\mathfrak{En}} \end{cases}. \tag{9.69}$$

For the speaker let $\bar{c}_\sigma := \min\{c(A) \,|\, A \in \mathscr{N}_\sigma \cap \mathscr{F}_{\mathfrak{En}}\}$, and:

$$\mathrm{Op}_\sigma^{\mathfrak{En}} := \{A \in \mathscr{N}_\sigma \cap \mathscr{F}_{\mathfrak{En}} \,|\, c(A) = \bar{c}_\sigma\}. \tag{9.70}$$

Then, $\bar{S}$ is defined by:

$$\bar{S}(A|\sigma) = \begin{cases} |\mathrm{Op}_\sigma^{\mathfrak{En}}|^{-1} & \text{if } A \in \mathrm{Op}_\sigma^{\mathfrak{En}} \\ 0 & \text{otherwise} \end{cases}. \tag{9.71}$$

We can show an equivalent to Lemma 2.3:

Lemma 9.5 *Let $\mathscr{S}$ be a set of support problems with canonical solution (S,H). Let $\mathscr{G}$ be a signalling game which supports $\mathscr{S}$. Let, furthermore, $\mathfrak{En}$ be an expected noise model which represents S. Then, the canonical solution $(\bar{S},\bar{H})$ always exists, and it is a Bayesian perfect equilibrium of $\mathscr{G}$. In addition, if we treat nominal costs as zero, then $(\bar{S},\bar{H})$ is Pareto dominating all other strategy pairs.*

This is the best result we can hope to achieve. We cannot exclude the possibility that there are signalling strategies which are more efficient than $(\bar{S},\bar{H})$.

For example, we may conceive an artificial signalling strategy for which the speaker says 'A' and snips with his fingers whenever he wants to say that there is a garage round the corner, and behaves exactly as if following $\bar{S}$ in all other situations. Then, this strategy is arguably more cost efficient than $\bar{S}$. As our framework does not exclude such artificial signalling strategies, we cannot in general prove that $(\bar{S},\bar{H})$ is Pareto dominating all other strategies.

9.2 Implikatures in EN models

We now consider the implicatures of an *ENmodel* $\mathfrak{En} = \langle \mathscr{S}, \mathscr{O}_\sigma, \mathscr{N}_\sigma \rangle$. As $\mathscr{O}_\sigma$ and $\mathscr{N}_\sigma$ are only there in order to represent strategies and their perturbations, they do not change the set of signalling games which support $\mathscr{S}$. As implicatures are an objective notion in our framework, and the objective level is described by signalling games, it follows that the implicatures of a signal F relative to s strategy S and an EN model $\mathfrak{En}$ are the same as its implicatures relative to S and $\mathscr{S}$. Hence, we set for $F \in \mathscr{F}$ for which $\exists \sigma \in \mathscr{S}\, S(F|\sigma) > 0$, and $R \subseteq \Omega$:

$$\langle \mathfrak{En}, S, H \rangle \models F +\!> R \iff \langle \mathscr{S}, S, H \rangle \models F +\!> R. \tag{9.72}$$

By Definition 5.4, this is equivalent to:

$$\langle \mathfrak{En}, S, H \rangle \models F +\!> R \iff \forall \mathscr{G} \in \mathrm{Supp}(\mathscr{S})\, \langle \mathscr{G}, S, H \rangle \models F +\!> R. \tag{9.73}$$

Here, $\mathrm{Supp}(\mathscr{S})$ is the set of all signalling games $\mathscr{G}$ which support $\mathscr{S}$.

Let $\mathfrak{En}_0$, $\mathfrak{En}_1$ be two EN models which represent the same strategy pair (S,H). Hence, $\mathscr{S}^{\mathfrak{En}_0} = \mathscr{S}^{\mathfrak{En}_1}$, and for each $\sigma \in \mathscr{S}^{\mathfrak{En}_i}$ it holds that $\mathscr{O}_\sigma^{\mathfrak{En}_i} = \{F \,|\, S(F|\sigma) > 0\}$. Then, Lemma 5.3 implies that:

$$\langle \mathfrak{En}_0, S, H \rangle \models F +\!> R \iff \langle \mathfrak{En}_1, S, H \rangle \models F +\!> R. \tag{9.74}$$

If $(\bar{S},\bar{H})$ is the *canonical solution* to $\mathfrak{En}$, we arrive with (5.43) at:

$$\left\langle \mathfrak{En}, \bar{S}, \bar{H} \right\rangle \models F +\!> R \iff \forall \sigma \in \mathscr{S}\, (F \in \mathrm{Op}_\sigma^{\mathfrak{En}} \Rightarrow P_S^\sigma(R) = 1). \tag{9.75}$$

Finally, we note a consequence of the definition for perturbed strategies S^ε. Let $\mathfrak{En} = \langle \mathscr{S}, \mathscr{O}_\sigma, \mathscr{N}_\sigma \rangle$ be an EN model which represents S and S^ε. Let H be an arbitrary hearer strategy. If F is such that $\exists \sigma\, S^\varepsilon(F|\sigma) > 0$ and $R \subseteq \Omega$, then we find again with (5.43) that:

$$\langle \mathfrak{En}, S, H \rangle \models F +\!> R \iff \forall \sigma \in \mathscr{S}\, (F \in \mathscr{O}_\sigma \Rightarrow P_S^\sigma(R) = 1), \tag{9.76}$$

and

$$\langle \mathfrak{En}, S^\varepsilon, H \rangle \models F +\!> R \iff \forall \sigma \in \mathscr{S}\, (F \in \mathscr{N}_\sigma \Rightarrow P_S^\sigma(R) = 1). \tag{9.77}$$

10 On the equilibrium properties of the canonical solution

Lemma 9.5 states that the best strategy pair that speaker and hearer can adopt in EN models is the *canonical* strategy pair defined in Section 9. More precisely, it states that the canonical strategy is a Bayesian perfect equilibrium, and, if we ignore nominal costs, it is even Pareto dominating all other strategies. The hearer's part of the canonical strategy is defined in (9.67), which in turn can be defined in terms of $H^A_{\mathbf{c}}$. The goal of this section is to prove Lemma 9.5. In addition, it contains some more fine grained characterisations of the canonical strategy.

Let $\mathscr{G} = \langle \Omega, \Theta, P, p, \mathscr{F}, \mathscr{A}, u \rangle$ be a signalling game which represents $\mathscr{S}$, and let $\mathfrak{En} = \langle \mathscr{S}, \mathscr{O}_\sigma, \mathscr{N}_\sigma \rangle$ be an EN model which represents a strategy S and the ε-approximations S^ε. Following our procedure in Section 8, we consider the following set:

$$Z^\varepsilon_H(A) := \{ \langle v, \sigma \rangle \mid P(v)\, p(\sigma|v)\, S^\varepsilon(A|\sigma) > 0 \wedge \exists a \notin \mathscr{B}(\sigma)\, H(a|A) > 0 \}. \quad (10.78)$$

For $\varepsilon = 0$, this corresponds to the set $Z(A)$ defined in (8.60). $Z^\varepsilon_H(A)$ is the set of all pairs of worlds v and support problems σ which have non–zero probability, for which the speaker answers A with non–zero probability, and for which H may choose a suboptimal action. In Proposition 8.2, we have shown that the hearer can improve if he reacts with a clarification request. Now we see that he can improve even more if he distinguishes between answers A which are elements of $\mathscr{F}_{\mathfrak{En}}$ and answers A which are not elements of $\mathscr{F}_{\mathfrak{En}}$. We first show how $\tilde{\mathscr{B}}(A)$ and $Z^\varepsilon_H(A)$ are related to each other:

Proposition 10.1 *Let $\mathscr{G}$ be a signalling game which fully supports the EN model $\mathfrak{En} = \langle \mathscr{S}, \mathscr{O}_\sigma, \mathscr{N}_\sigma \rangle$. Assume that $\mathfrak{En}$ represents S and the ε-approximations S^ε. Let $A \in \mathscr{N}$. Then:*

$$\tilde{\mathscr{B}}(A) = \emptyset \Leftrightarrow \forall H\, (H(\mathbf{c}|A) < 1 \Rightarrow Z^\varepsilon_H(A) \neq \emptyset). \quad (10.79)$$

Proof: We first prove "⇒": Assume that $\tilde{\mathscr{B}}(A) = \emptyset$. Then, by definition:

$$\forall a \in \mathscr{A} \setminus \{\mathbf{c}\}\, \exists \sigma, \sigma' : A \in \mathscr{N}_\sigma \cap \mathscr{N}(\sigma') \wedge a \notin \mathscr{B}(K_\sigma) \cap \mathscr{B}(K_{\sigma'}).$$

As P^σ_S is fully reliable for each $\sigma \in \mathscr{S}$, it follows that $\mathscr{B}(K_\sigma) = \mathscr{B}(\sigma)$ and $\mathscr{B}(K_{\sigma'}) = \mathscr{B}(\sigma')$. Let $a \in \mathscr{A} \setminus \{\mathbf{c}\}$, and let H be any hearer strategy with $H(a|A) > 0$. Then, there exists σ such that $A \in \mathscr{N}_\sigma$ and $a \notin \mathscr{B}(\sigma)$. As $\mathfrak{En}$ is supported by $\mathscr{G}$ and S^ε represented by $\mathfrak{En}$, it follows that $P(v)\, p(\sigma|v)\, S^\varepsilon(A|\sigma) > 0$; hence we can find a v such that $\langle v, \sigma \rangle \in Z^\varepsilon_H(A)$ and an a with $H(a|A) > 0 \wedge a \notin \mathscr{B}(\sigma)$. Therefore $Z^\varepsilon_H(A) \neq \emptyset$.

"⇐": Assume that $\tilde{\mathscr{B}}(A) \neq \emptyset$. Let $a \in \tilde{\mathscr{B}}(A)$ and set $H(a|A) = 1$. Assume that $P(v)\, p(\sigma|v)\, S^\varepsilon(A|\sigma) > 0$. As P^σ_S is fully reliable for each $\sigma \in \mathscr{S}$, it follows

by definition of $\tilde{\mathscr{B}}(A)$ that $a \in \mathscr{B}(\sigma)$. Hence, $\langle v, \sigma \rangle \notin Z_H^\varepsilon(A)$. As v and σ are arbitrary, it follows that $Z_H^\varepsilon(A) = \emptyset$. ■

This shows how we can improve over Proposition 8.2: Only if $\mathscr{B}(A) = \emptyset$ the hearer reacts with a clarification request, else he chooses an act from $\mathscr{B}(A)$. $\mathscr{B}(A) = \emptyset$ is equivalent to $A \in \mathscr{F}_{\mathfrak{En}}$. The next proposition tells us how expected utilities behave if the hearer changes his strategy from $H(\,.\,|A)$ to H_X^A.

Proposition 10.2 *Let $\mathscr{G}$ be a signalling game which fully supports the EN model $\mathfrak{En} = \langle \mathscr{S}, \mathscr{O}_\sigma, \mathscr{N}_\sigma \rangle$. Assume that $\mathfrak{En}$ represents S and the ε-approximations S^ε. Let $A \in \mathscr{N}$. Let $\mathscr{E}_H$ be the hearer's expected utility if the speaker follows strategy S, $\mathscr{E}_H^\varepsilon$ be the hearer's expected utility if the speaker follows strategy S^ε, and $\mathscr{E}_H^\sim$ be the hearer's expected utility if the speaker follows some other strategy $\tilde{S}$. All the expected utilities are defined relative to the probabilities in $\mathscr{G}$. Let $M_\sigma = \max_{a \in \mathscr{A} \setminus \{\mathbf{c}\}} \mathscr{E}_S(a|\sigma)$, and let μ be as in (6.52). In the following equations, let $X \subseteq \mathscr{A}$ be such that $\mathbf{c} \notin X$. Then:*

1. $\mathscr{E}_H(H|A) = \mathscr{E}_H(H_X^A|A) = \sum_\sigma \mu(\sigma|A)\, M_\sigma - c(A)$ *for* $X \subseteq \tilde{\mathscr{B}}(A)$,
2. $\mathscr{E}_H^\varepsilon(H_X^A|A) < \mathscr{E}_H^\varepsilon(H^A|A) = \sum_\sigma \mu^\varepsilon(\sigma|A)\, M_\sigma - c(A)$ *for* $X \setminus \tilde{\mathscr{B}}(A) \neq \emptyset$,
3. *Let $\tilde{S}$ be given with $\tilde{S}(\mathscr{F}_{\mathfrak{En}}|\sigma) = 1$ for all σ, and let $X \setminus \tilde{\mathscr{B}}(A) \neq \emptyset$, then*

$$\mathscr{E}_H^\sim(H_X^A|A) < \mathscr{E}_H^\sim(H^A|A) = \sum_\sigma \mu(\sigma|A)\, M_\sigma - c(A).$$

Proof: We first show that $\mathscr{E}_H(H|A) = \mathscr{E}_H(H_X^A|A)$ for $X \subseteq \tilde{\mathscr{B}}(A)$: By (6.53) $\mathscr{E}_H(H|A) = \sum_\sigma \mu(\sigma|A) \sum_a H(a|A)\, \mathscr{E}_S(a|\sigma) - c(A)$; from (6.57) it follows that $H(a|A) > 0 \Rightarrow \mathscr{E}_S(a|\sigma) = M_\sigma$; hence $\mathscr{E}_H(H|A) = \sum_\sigma \mu(\sigma|A)\, M_\sigma - c(A)$; as $X \subseteq \tilde{\mathscr{B}}(A)$, it follows again with (6.53) and (6.57) that

$$\mathscr{E}_H(H_X^A|A) = \sum_\sigma \mu(\sigma|A) \sum_{a \in X} H_X^A(a|A)\, \mathscr{E}_S(a|\sigma) - c(A) = \sum_\sigma \mu(\sigma|A)\, M_\sigma - c(A).$$

Next, we turn to $\mathscr{E}_H^\varepsilon(H_X^A|A) < \mathscr{E}_H^\varepsilon(H^A|A)$ for all X with $\tilde{\mathscr{B}}(A) \subsetneqq X$: By (6.53), $\mathscr{E}_H^\varepsilon(H_X^A|A) = \sum_\sigma \mu^\varepsilon(\sigma|A) \sum_a H_X^A(a|A)\, \mathscr{E}_S(a|\sigma) - c(A)$; we can divide $\sum_\sigma$ into the sum over the set $M_0 = \{\sigma \mid H_X^A(\mathscr{B}(\sigma)|A) = 1\}$ plus the sum over the set $M_1 = \{\sigma \mid H_X^A(\mathscr{B}(\sigma)|A) < 1\}$; as $\tilde{\mathscr{B}}(A) \subsetneqq X$, the second set is not empty; as $H^A(a|A) > 0 \Rightarrow a \in \tilde{\mathscr{B}}(A)$, it follows for M_1 that

$$\sum_{M_1} \mu^\varepsilon(\sigma|A) \sum_a H_X^A(a|A)\, \mathscr{E}_S(a|\sigma) - c(A) < \sum_{M_1} \mu^\varepsilon(\sigma|A)\, M_\sigma - c(A) = \\ = \sum_{M_1} \mu^\varepsilon(\sigma|A) \sum_a H^A(a|A)\, \mathscr{E}_S(a|\sigma) - c(A).$$

By (6.57) it follows that equality holds if we replace M_1 by M_0. This proves the claim.

Finally, the proof of $\mathscr{E}_H^{\sim}(H_X^A|A) < \mathscr{E}_H^{\sim}(H^A|A) = \sum_\sigma \mu(\sigma|A)\, M_\sigma - c(A)$ is almost identical to the previous case. ■

In order to improve the readability of formulas, we use the abbreviation $NC(s)$ for expressing the fact that a speaker strategy s is not optimal but differs from an optimal strategy s' only by a positive term with nominal costs; in addition, these nominal costs can only be reduced by a change of the speaker strategy s, not by a change of the hearer strategy. That is, if we write $EU(s'|h) = EU(s|h) - NC(s)$, we mean that $EU(s'|h) < EU(s|h)$ such that first $EU(s'|h) - EU(s|h)$ is nominal, and second there is no strategy h' for which $EU(s|h') > EU(s|h)$. We find:

Proposition 10.3 *Let $\mathscr{S}$ be a set of support problems with canonical solution (S,H). Let $\mathscr{G}$ be a signalling game which supports $\mathscr{S}$. Let, furthermore, $\mathfrak{En} = \langle \mathscr{S}, \mathscr{O}_\sigma, \mathscr{N}_\sigma \rangle$ be an expected noise model which represents S, and let $(\bar{S},\bar{H})$ be its canonical solution. Then:*

1. $EU(\bar{S}|\bar{H}) = \sum_\sigma \mu_\Theta(\sigma)\,(M_\sigma - \sum_A \bar{S}(A|\sigma)\, c(A))$;
2. $EU(S^\varepsilon|\bar{H}) = EU(\bar{S}|\bar{H}) - NC(S^\varepsilon)$.

If for all $\sigma \in \mathscr{S}$ $\mathscr{O}_\sigma \subseteq \mathscr{F}_{\mathfrak{En}}$, then

3. $EU(S|H) = EU(S,\bar{H}) = EU(\bar{S}|\bar{H}) - NC(S)$.

Furthermore, if $\tilde{S}$ is such that $\exists\sigma\, \tilde{S}(\mathscr{F}_{\mathfrak{En}}|\sigma) < 1$, then

4. $EU(\tilde{S},\bar{H}) = EU(\bar{S}|\bar{H}) - NC(\tilde{S})$

Proof: 1) By definition, $\bar{S}(A|\sigma) > 0$ implies $\bar{H}(a|A) > 0 \Rightarrow a \in \mathscr{B}(\sigma)$. Then, (6.55) and (2.1) imply that $EU(\bar{S}|\bar{H}) = \sum_v P(v) \sum_\sigma p(\sigma|v) \sum_A \bar{S}(A|\sigma)\,(M_\sigma - c(A))$, which equals $\sum_\sigma \mu_\Theta(\sigma)\,(M_\sigma - \sum_A \bar{S}(A|\sigma)\, c(A))$.

2) Let $\mu^\varepsilon(A) := \sum_v P(v) \sum_\sigma p(\sigma|v)\, S^\varepsilon(A|\sigma)$. If $\mu^\varepsilon(A|\sigma) > 0$ and $\tilde{\mathscr{B}}(A) = \emptyset$, then $\bar{H}(\mathbf{c}|A) = 1$, i.e. the hearer will react to A with a clarification request $\mathbf{c}$. If $\mu^\varepsilon(A|\sigma) > 0$ and $\tilde{\mathscr{B}}(A) \neq \emptyset$, then S^ε will produce higher costs than $\bar{S}$, iff $c(A)$ is more costly than $\bar{c}_\sigma$. Hence, with $\mu_\Theta(\sigma) = \sum_v P(v)\, p(\sigma|v)$, we arrive at:

$$\begin{aligned} EU(S^\varepsilon|\bar{H}) &= EU(\bar{S}|\bar{H}) - \Big(c(\mathbf{c})\, \mu^\varepsilon(\{A \mid \tilde{\mathscr{B}}(A) = \emptyset\}) + \\ &\qquad + \sum_\sigma \mu_\Theta(\sigma) \sum_A S^\varepsilon(A|\sigma)\,(c(A) - \bar{c}_\sigma)\Big) \\ &= EU(\bar{S}|\bar{H}) - NC(S^\varepsilon). \end{aligned} \tag{10.80}$$

This proves the second claim.

3) The first equation is trivially true. The second equation follows similarly to 2) as $S(A|\sigma) > 0$ implies $c(A) \geq \bar{c}_\sigma$; hence:

$$EU(S|\bar{H}) = EU(\bar{S}|\bar{H}) - \sum_\sigma \mu(\sigma) \sum_A S(A|\sigma)\,(c(A) - \bar{c}_\sigma). \tag{10.81}$$

This again proves the claim.

4) Assume that $\exists\sigma\,\tilde{S}(\mathscr{F}_{\mathfrak{En}}|\sigma) < 1$. We split $\mathscr{S}$ into $M_0 = \{\sigma \mid \tilde{S}(\mathscr{F}_{\mathfrak{En}}|\sigma) = 1\}$ and $M_1 = \{\sigma \mid \tilde{S}(\mathscr{F}_{\mathfrak{En}}|\sigma) < 1\}$. Then, for each $\sigma \in M_1$, we split $\mathscr{F}$ into $F_0^\sigma = \{A \in \mathscr{F}_{\mathfrak{En}} \mid \tilde{S}(A|\sigma) > 0)\}$ and $F_0^\sigma = \{A \in \mathscr{F} \setminus \mathscr{F}_{\mathfrak{En}} \mid \tilde{S}(A|\sigma) > 0)\}$. As $\exists\sigma\,\tilde{S}(\mathscr{F}_{\mathfrak{En}}|\sigma) < 1$, it follows that $\exists\sigma\,\tilde{S}(F_0^\sigma|\sigma) > 0$. Then, let $\tilde{S}'$ be the strategy which results from replacing each $A \in F_0^\sigma$ by a $B \in \mathscr{N}_\sigma \cap \mathscr{F}_{\mathfrak{En}}$. Then, clearly, $EU(\tilde{S}|\bar{H}) < EU(\tilde{S}'|\bar{H})$, and by definition $EU(\tilde{S}'|\bar{H}) = EU(\bar{S}|\bar{H}) - NC(\tilde{S}')$. As the difference between $EU(\tilde{S}|\bar{H})$ and $EU(\tilde{S}'|\bar{H})$ is only nominal, it also follows that $EU(\tilde{S}|\bar{H}) = EU(\bar{S}|\bar{H}) - NC(\tilde{S})$. ■

With these preparations, we can finally show:

Proof of Lemma 9.5: From the third claim of Prop. 10.2, it follows that the hearer has no better strategy against $\bar{S}$ than $\bar{H}$, in particular, $\bar{H}$ satisfies the Bayesian condition. From the fourth claim of Prop. 10.3 it follows that the speaker prefers strategies $\tilde{S}$ with $\tilde{S}(\mathscr{F}_{\mathfrak{En}}|\sigma) = 1$ over strategies $\tilde{S}$ with $\tilde{S}(\mathscr{F}_{\mathfrak{En}}|\sigma) < 1$. By definition, strategies $\tilde{S}$ which satisfy $\tilde{S}(\mathscr{F}_{\mathfrak{En}}|\sigma) = 1$ cannot be better than $\bar{S}$ against $\bar{H}$. Hence, it follows that $(\bar{S},\bar{H})$ is a Bayesian perfect equilibrium of $\mathscr{G}$. From the first claim of Prop. 10.3, it immediately follows that $(\bar{S},\bar{H})$ is Pareto dominating all other strategy pairs if nominal costs are treated as zero. ■

11 Nominality

In this section we supplement a precise definition of *nominal* costs. As the technical details are of minor interest to the purposes of the present paper, we present only the bare essentials.

Definition 11.1 (Nominality) *Let u be a function which takes arguments $\mathbf{a} = \langle a_1,\ldots,a_n\rangle$. Let u_1 and u_2 be two functions for which $u(\mathbf{a}) = u_1(\mathbf{a}) + u_2(\mathbf{a})$. By saying that u_2 is* nominal *with respect to u, we say that for any continuous function f and arguments $\mathbf{a},\mathbf{b}$ the inequality $f(u(\mathbf{a})) \leq f(u(\mathbf{b}))$ means that*

$$\lim_{k\to 0^+} \operatorname{sgn}\big(f(u_1(\mathbf{b}) + k\,u_2(\mathbf{b})) - f(u_1(\mathbf{a}) + k\,u_2(\mathbf{a}))\big) \geq 0. \tag{11.82}$$

In this formula, $k \to 0^+$ means that we only consider sequences of $k > 0$ which converge to 0. The signum function sgn is defined as follows:

$$\operatorname{sgn}(x) := \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0 \end{cases}.$$

One should keep in mind that there are continuous f for which the limit in (11.82) is not defined. But it is always defined for constant, linear, or monotonic continuous functions f. For linear functions, (11.82) is equivalent to:

$$f(u_1(\mathbf{a})) < f(u_1(\mathbf{b})) \vee (f(u_1(\mathbf{a})) = f(u_1(\mathbf{b})) \wedge f(u_2(\mathbf{a})) < f(u_2(\mathbf{b}))). \tag{11.83}$$

As example, we consider the speaker's expected utility of uttering F given a hearer strategy H for a support problem $\langle \Omega, P_S, P_H, \mathscr{F}, \mathscr{A}, u, c, [\![.]\!] \rangle$. The utility function u is of the form $u(v,F,a) = u(v,a) + c(F)$ with nominal c. Hence, we set $u_1(a,F,a) = u(v,a)$ and $u_2(a,F,a) = -c(F)$. The speaker's expected utility is defined as:

$$EU_S(F) = \sum_{v \in \Omega} P_S(v) \sum_{a \in \mathscr{B}(F)} H(a|F)\, u(v,F,a). \tag{11.84}$$

As EU_S is linear, we find:

$$EU_S^k(F) := EU_S^0(F) - k\, c(F), \tag{11.85}$$

Hence, the nominality of c entails that for all $F_0, F_1 \in \mathscr{F}$:

$$EU_S(F_0) \leq EU_S(F_1) \;:\Leftrightarrow\; \lim_{k \to 0^+} \operatorname{sgn}\Big(EU_S^k(F_1) - EU_S^k(F_0)\Big) \geq 0. \tag{11.86}$$

Clearly, it follows that

$$EU_S^0(F_0) < EU_S^0(F_1) \Rightarrow EU_S(F_0) < EU_S(F_1). \tag{11.87}$$

For clarification requests we have to generalise the definition slightly. Expected utilities with clarification requests divide into a non–nominal term which depends on the speakers signal A, and a nominal term which is the sum of the costs for uttering A and the costs due to the clarification request $\mathbf{c}$. Hence, we generalise (11.86) as follows: For finite sets $X \subseteq \operatorname{dom} \mathbf{c}$ we set

$$c(X) := \sum_{F \in X} c(F), \text{ and } EU_S^k(F_i, X) = EU_S^k(F_i) - k\, c(X). \tag{11.88}$$

Then, $EU_S(F_0,X_0) \leq EU_S(F_1,X_1)$ iff:

$$\lim_{k \to 0^+} \operatorname{sgn}\Big(EU_S^k(F_1,X_1) - EU_S^k(F_0,X_0)\Big) \geq 0. \tag{11.89}$$

For example, F_0 and F_1 may be two possible utterances, and $X_0 = \{\mathbf{c}\}$ and $X_1 = \emptyset$. It follows by definition that

$$EU_S^0(F_0) < EU_S^0(F_1) \Rightarrow EU_S(F_0,X_0) < EU_S(F_1,X_1). \tag{11.90}$$

References

Benz, A. (2006). Utility and Relevance of Answers. In Benz, A., Jäger, G., and van Rooij, R., editors, *Game Theory and Pragmatics*, pages 195–214. Palgrave Macmillan, Basingstoke.

Benz, A. (2007). On Relevance Scale Approaches. In Puig-Waldmller, E., editor, *Proceedings of the Sinn und Bedeutung 11*, pages 91–105.

Benz, A. (2008). How to Set Up Normal Optimal Answer Models. Ms, ZAS, Berlin.

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

Franke, M. (2009). *Signal to Act: Game Theory in Pragmatics*. PhD thesis, Universiteit van Amsterdam.

Grice, H. P. (1957). Meaning. *Philosophical Review*, 66:377–388.

Grice, H. P. (1989). *Studies in the Way of Words*. Harvard University Press, Cambridge MA.

Groenendijk, J. and Stockhof, M. (1991). Dynamic predicate logic. *Linguistics & Philosophy*, 14:39–100.

Jäger, G. and Ebert, C. (2009). Pragmatic Rationalizability. In Riester, A. and Solstad, T., editors, *Proceedings of Sinn und Bedeutung*, volume 13.

Lewis, D. (2002). *Convention*. Blackwell Publishers, Oxford. First published by Harvard University Press 1969.

Myerson, R. (1978). Refinements of the Nash equilibrium concept. *International Journal of game theory*, 7:73–80.

Parikh, P. (1990). Situations, Games, and Ambiguity. In Cooper, R., Mukai, K., and Perry, J., editors, *Situation Theory and its Applications*, volume 1, pages 449–469. CSLI Lecture Notes, Stanford, CA.

Parikh, P. (2001). *The Use of Language*. CSLI Publications, Stanford.

Parikh, P. (2006). Pragmatics and Games of Partial Information. In Benz, A., Jäger, G., and van Rooij, R., editors, *Game Theory and Pragmatics*, pages 101–121. Palgrave Macmillan, Basingstoke.

Pearle, J. (2000). *Causality - Models, Reasoning, and Inference*. Cambridge University Press, Cambridge.

Savage, L. (1972). *The Foundations of Statistics*. Dover Publications, New York. revised and enlarged version; first published by John Wiley & Sons, 1954.

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