R.A.M.vanRooij@uva.nl

Comparing Questions and Answers: A bit of Logic, a bit of Language, and some bits of Information.*

Robert van Rooij

Institute for Logic, Language and Computation
University of Amsterdam

Abstract. In this paper various ways of measuring the value of questions and answers are discussed and compared. For measurement, Information Theory and Decision Theory are used. The analysis is based on the assumptions that answers denote propositions denoted by sets of possible worlds, while questions denote partitions.

1 Introduction

Notions like 'entropy' and '(expected) value of observations' are widely used in science to determine which experiment to conduct to make a better informed choice between a set of scientific theories that are all consistent with the data. But these notions seem to be almost equally important for our use of language in daily life as they are for scientific inquiries.

I will make use of these notions to measure how 'good' particular questions and answers are in particular circumstances. In doing so, I will extend and/or refine the *qualitative* approach towards such measurements proposed by Groenendijk & Stokhof (1984). The refinements are due to the fact that I also take into account (i) probabilities, (ii) utilities, and (iii) the idea that we ask questions to resolve decision problems.

In this paper I will first explain Groenendijk & Stokhof's partition based analysis of questions, and then discuss their *qualitative* method of measurement. Next, I will take also *probabilities* into account, and show how a natural *quantitative* measure of informativity can be defined in terms of it. Following the lead of Communication Theory and Inductive Logic, I will then show that we can also describe a natural measure of the informative value of questions and answers in terms of *conditional entropy*, when we take into account that questions

* I appreciate it a lot that Giovanni Sommaruga invited me to submit this paper to the present volume, given that the bulk of it was written already in 2000. I would like to thank the following people for discussion and comments: Alexandru Baltag, Balder ten Cate, Paul Dekker, Roberto Festa, Jeroen Groenendijk, Emiel Krahmer, Marie Nilsenova, and Yde Vennema. Since the time that I wrote this paper, I have published two articles (van Rooij 2004a,b) that could be thought of as successors of this paper.

are asked to resolve *decision problems*. Finally, I will argue that to measure the value of questions and answers we should in general also take *utilities* seriously, and following standard practice in Statistical Decision Theory, I show how some intuitively appealing utility values can be specified.

2 The semantics of declaratives and interrogatives

The perhaps most 'natural' conception of 'meaning', at least in its point of departure, identifies 'meaning' with *naming*. The meaning of an expression is that what the expression *refers to*, or *is about*. What meaning does is to establish a *correspondence* between expressions in a *language* and things in the (model of the) *world*. For simple expressions like proper names and simple declarative sentences, this view of meaning is natural and simple. The meaning of *John* is the object it refers to, while the meaning of a simple declarative sentence like *John is sick* could then be the *fact* that John is sick. Beyond this point of departure, things are perhaps less natural. What, for example, should be the things out in the world that a negated sentence like *John is not sick* is about, and what should *John is sick* be about if the sentence is false? Notice that to be a competent speaker of English one has to know what it means for *John is sick* to be true or false. So a minimal requirement for any theory of meaning would be that one knows the meaning of a declarative sentence if one knows under which circumstances it is, or would be, true. The proposal of formal semanticists to solve our above conceptual problems is to stick to this minimal requirement: identify the meaning of a declarative sentence with the *conditions*, or *circumstances* under which the sentence is *true*. These circumstances can, in turn, be thought of as the ways the world might have been, or *possible worlds*. Thus, the meaning of a sentence can be thought of as the set of circumstances, or possible worlds, in which it is true. This latter set is known in possible worlds semantics as the *proposition* expressed by the sentence. We will denote the meaning of any declarative sentence A by $[[A]]$, and identify it with the set of worlds in which A is true (where W is the set of all possible worlds):[1]

$$[[A]] \quad = \quad \{w \in W : \ A \text{ is true in } w\}.$$

Just as it is standard to assume that you know the meaning of a declarative sentence when you know under which circumstances this sentence is true, Hamblin (1958) argues that you know the meaning of a question when you know what counts as an appropriate answer to the question. Because we answer a question by making a statement that expresses a proposition, this means that the meaning of a question as linguistic object (interrogative sentence) can be equated with the set of propositions that would be expressed by the appropriate linguistic answers. This gives rise to the problem what an appropriate linguistic answer is to a question.

[1] Here, and elsewhere in this paper, I will assume that we analyze sentences with respect to a fixed intensional model.

For a *yes/no*-question like *Is John sick?* it is widely agreed that it has only one appropriate true answer; *Yes* in case John is sick, and *No* when John is not sick. This means that with respect to each world a *yes/no*-question simply expresses a proposition; the proposition expressed by the true appropriate answer in that world. If we represent a *yes/no*-question simply by a formula like $?A$, where A is a sentence, and assume that $[[A]]^w$ denotes the truth value of A in w, the proposition expressed by question $?A$ in world w is:

$$[[?A]]^w \quad = \quad \{v \in W : \ [[A]]^v = [[A]]^w\}.$$

Given this analysis of polar interrogative sentences, the question arises what the meaning of a *wh*-question is; i.e. what counts in a world as an appropriate true answer to questions like *Who is sick?* and *Who kissed whom?*

Groenendijk & Stokhof (1984) have argued on the basis of linguistic phenomena that not only *yes/no*-questions, but also (multiple) *wh*-questions can in each world only have *one true (complete) answer*. They argue that for John to know the answer to the question *Who is sick?*, for instance, John must know of *each* (relevant) individual *whether* he or she is sick.[2]

Representing questions abstractly by $?P$, where P is an n-ary predicate, John gives in w the true and complete answer to the above question just in case he gives an answer that entails the following proposition, where $[[P]]^v$ denotes the extension of predicate P in world v:

$$[[?P]]_g^w \quad = \quad \{v \in W | \ [[P]]^v \ = \ [[P]]^w \ \}.$$

If P is a 1-ary predicate like *is sick*, $[[P]]^w$ denotes the set of individuals that are sick in w, and $[[?P]]^w$ denotes the set of worlds where P has the same extension as in world w. If P is a binary predicate like *kissed*, $[[P]]^w$ denotes the set of ordered pairs $\langle d, d' \rangle$ such that d kissed d' in w, and $[[?P]]^w$ denotes the set of worlds where the same individuals kissed each other as in world w. An interesting special case is when P is a zero-ary predicate, i.e., when P is a sentence and when the question is thus a *yes/no*-question. In that case the proposition expressed by the question in a world will be exactly the same as the proposition determined via our second equation. Thus, according to Groenendijk & Stokhof's (1982) proposal, we should not only treat single and multiple *wh*-questions in the same way, but we should analyze *yes/no*-questions in a similar way, too.

Suppose, contrary to Hamblin's suggestion, that we equate the meaning of a question with the meaning of its true answer. This would immediately allow us to define an entailment relation between questions.[3] We can just say that one question entails another, just in case the proposition expressed by the true

[2] This doesn't mean that everybody agrees. For a discussion of some problems, and alternative analyses of questions, see Groenendijk & Stokhof (1997).

[3] In this paper I will use the term 'question' not only for interrogative sentences, but also for the meanings they express. Something similar holds for the term 'answer'. I hope this will never lead to confusion.

answer to the first question entails the proposition expressed by the true answer to the second question. And given an entailment relation between questions, it seems only natural to say that one question is 'better', or 'more informative' than another exactly when the former question entails the latter.

However, the above suggested entailment-relation between questions, and the thus induced 'better than'-relation, doesn't seem to be very natural. Suppose that in fact both John and Mary are sick. In that case it holds that the true answer to the question *Are John and Mary sick?* entails the true answer to the question *Is John sick?*, and thus it is predicted that the first question also entails the second. But this prediction seems to be wrong; the first question does intuitively not entail the second question because when Mary were in fact not sick (although John still is), the true answer to the first question would no longer entail the true answer to the second question. What this suggests is that the entailment-relation between questions does not just depend on how the world *actually* is, but also on how the world *could have been*.

Above, we have defined the proposition expressed by a question with respect to the real world, w. The above discussion suggests that to define an entailment relation between propositions, we should *abstract* away from how the actual world looks like. We should say that one question entails another just in case knowing the true answer to the former means that you also know the true answer to the latter, *however the world looks like*. Thus, $?P_1$ entails $?P_2$, $?P_1 \models ?P_2$, just in case it holds for *every* world w that $[[?P_1]]^w$ is a subset of $[[?P_2]]^w$:

$$?P_1 \models ?P_2 \quad \text{iff} \quad \forall w : [[?P_1]]^w \subseteq [[?P_2]]^w.$$

We might also define this entailment relation between questions more directly in terms of their meanings. In order to do this, we should think of the meaning of a question itself no longer simply as a proposition, but rather as a *function* from worlds to propositions (answers):

$$[[?P]] \quad = \quad \lambda w.\{v \in W \mid [[P]]^v \ = \ [[P]]^w \ \}.$$

Notice that this function from worlds to propositions is simply equivalent to the following *set* of propositions:

$$\{\{v \in W \mid [[P]]^v \ = \ [[P]]^w \ \} \mid w \in W\}.$$

and, due to the assumption that a question has in each world only one true answer, this set of propositions *partitions* the set of worlds W. A partition of W is a set of mutually exclusive non-empty subsets of W such that their union equals W. In fact, the partition that is induced in this way by a question is exactly what Groenendijk & Stokhof (1984) have proposed to call the *semantic meaning*, or *intension*, of a question, and they distinguish it from the *extension* a question has, $[[?P]]^w = [[?P]](w)$, in the particular world w. Notice that Groenendijk & Stokhof's account is in accordance with Hamblin's proposal: the meaning of a question is represented by its set of possible appropriate answers.

We have seen that the partition semantics of questions is based on the assumption that every question can in each world have at most *one* semantic answer. Thus, if you ask me *Who of John and Mary are sick?*, I can only resolve the question according to this analysis by giving an *exhaustive* answer where I tell for both John and Mary whether they are sick or not. It might, however, be the case that I only know whether John is sick, and that I just respond by saying *(At least) John is sick*. This response will obviously not resolve the whole issue, and thus will not count as a *complete*, or semantic, answer to the question. Still, it does count as an answer to the question, although only a *partial* one. We can say that an assertion counts as a partial answer to the question iff it is a non-contradictory proposition that is incompatible with at least one cell of the partition induced by the question. In our above example, for instance, the response *(At least) John is sick* counts as a partial answer to the question, because it is incompatible with 2 of the 4 cells of the partition induced by the question. Observe that according to our characterization of partial answerhood, it holds that a complete, semantic, answer to the question also is incompatible with at least one cell of the partition, and thus also counts as a partial answer. So we wee that some partial answers are more informative, and better, than others.

Suppose that Q and Q' are two partitions of the logical space that are induced by two interrogative sentences. Let us also assume for simplicity that we can equate the meaning of an interrogative sentence with the question itself. Making use of the fact that every question has according to their semantics (at most) one answer in each world, Groenendijk & Stokhof (1984) can define the entailment-relation between questions directly in terms of their meanings making use of a generalized subset-relation, '$\sqsubseteq$', between partitions. Remember that according to our above requirement, for question Q to *entail* question Q', $Q \models Q'$, it must be the case that knowing the true answer to Q means that you also know the true answer to Q', *however the world looks like*. In terms of Groenendijk & Stokhof's (1984) partition semantics this comes down to the natural requirement that for every element of Q there must be an element of Q' such that the former entails the latter, i.e. $Q \sqsubseteq Q'$:

$$Q \models Q' \quad \text{iff} \quad Q \sqsubseteq Q' \quad \text{iff} \quad \forall q \in Q : \exists q' \in Q' : q \subseteq q'.$$

According to this definition it follows, for instance, that the *wh*-question *Who of John and Mary are sick?* entails the *yes/no*-question *Is John sick?*, because every (complete) answer to the first question entails an answer to the second question. And indeed, when you know the answer to the first question, the second question can no longer be an issue. Something similar is the case for the multiple *wh*-question *Who kissed whom?* and the single *wh*-question *Who kissed Mary?*; learning the answer to the first question is *more informative* than learning the answer to the second question.

3 Comparing questions and answers qualitatively

3.1 A semantic comparison

If somebody asks you who murdered Smith, he would not be satisfied with an answer like *The murderer of Smith*. Although this answer will obviously be true, it is unsatisfactory because the answer will *not* be *informative*. Indeed, it is generally agreed that in normal circumstances the utterance of an interrogative sentence is meant as a means to acquire information.

If the aim of the question is to get some information, it seems natural to say that Q is a *better* question than Q', if it holds that whatever the world is, knowing the true answer to question Q means that you also know the true answer to Q', i.e. $Q \sqsubseteq Q'$. As we have seen above, this would mean that the question *Who of John and Mary are sick?* is a 'better' question than *Is John sick?*, because the former question entails the latter. Notice that by adopting this approach, the *value*, or *goodness*, of a question is ultimately reduced to the pure *informativity* of the expected answer.

Not only can we compare questions to each other with respect to their 'goodness', the same can be done for *answers*. We have seen in the previous section that complete answers are special kinds of partial answers; the most informative partial answers that are true in the worlds of just one cell of a partition. This suggests, perhaps, the following proposal; say that one answer is 'better' than another, just in case the former *entails* the latter. But this would be mistaken, for it would wrongly predict that we prefer *over*informative answers to answers that are just complete. If I ask you, for instance, whether John is sick, I would be very puzzled by your answer *Yes, John is sick, and it is warm in Africa*. The second conjunct to the answer seems to be completely *irrelevant* to the issue, and thus should not be mentioned.

So it seems that we should measure the 'goodness' of an answer mostly in terms of the partition induced by the question. And indeed, this is exactly what Groenendijk & Stokhof (1984) propose. Define A_Q as the set of cells of partition Q that are compatible with answer A:[4]

$$A_Q \quad = \quad \{q \in Q : \; q \cap A \neq \emptyset\}.$$

Notice now that one partial answer can be more informative than another one because it is incompatible with more cells of the partition than the other one. Remember that the answer $A =$ *(At least) John is sick* counts as a partial answer to the question $Q =$ *Who of John and Mary are sick?*, and is incompatible with 2 of the 4 cells of the partition. The answer $B =$ *If Mary is not sick, then neither is John* also counts as a partial answer to the question, because it is incompatible with 1 cell of the partition. But it is a weaker answer than *(At least) John is*, because it is entailed by the latter and incompatible with less cells of the partition than the former one, i.e. $A_Q \subset B_Q$. Groenendijk & Stokhof

[4] From now on I tend to use the same notation both for a declarative sentence and the proposition it expresses. I hope this will never lead to confusion.

propose that when answer A is incompatible with more cells of the partition than answer B, i.e. $A_Q \subset B_Q$, the former should be counted as a better answer to the question than the latter.

But what if two answers are incompatible with the same cells of the partition, i.e. if $A_Q = B_Q$? It is possible that when two partial answers to a question that are incompatible with, for example, just one cell of the partition, one of them can be more informative than the other because the former *entails* the latter. In our above example, for instance, not only *(At least) John is sick*, but also *John is sick, and it is warm in Africa* is an answer that is incompatible with just two cells of the partition induced by the question. As we have suggested above already, the former counts in that case as a better answer than the latter, because it doesn't give extra *irrelevant* information. Thus, in case $A_Q = B_Q$, A is a better answer than B iff $A \supset B$.

Combining both constraints, Groenendijk & Stokhof (1984) propose that A is (quantitatively) a *better* semantic answer to question Q than B, $A >_Q B$, by defining the latter notion as follows:

$$A >_Q B \quad \text{iff either (i)} \quad A_Q \subset B_Q, \text{ or}$$
$$\text{(ii)} \quad A_Q = B_Q, \text{ and } A \supset B.$$

Lewis (1988) and Groenendijk (1999) defined a notion of *aboutness* in terms of which answers can be compared in a more direct way.[5] They say that answer A is *about* question Q just in case the following condition is satisfied:

$$A \text{ is } \textit{about } Q \quad \text{iff} \quad \forall q \in Q : q \subseteq A \text{ or } q \cap A \neq \emptyset.$$

Thus, when A is true/false in a world w, it should be the case that A is also true/false in any world v that is an element of the same cell of the partition Q as w is. Notice that because Q is a partition, the above definition of aboutness is equivalent to the following condition:

$$A \text{ is } \textit{about } Q \quad \text{iff} \quad \bigcup A_Q = A.$$

This notion of aboutness intuitively corresponds with the second condition in the definition of $>_Q$ that no extra *irrelevant* information should be given. Using the standard Stalnakerian (1978) assertion conditions, we might say that with respect to a certain question, an assertion is *relevant* if it is (i) consistent, (ii) informative, and (iii) is about the question. In terms of such a notion of relevance, we can re-define the above 'better than' relation, $A >_Q B$, between relevant answers A and B to question Q simply as follows:

$$A >_Q B \quad \text{iff} \quad A \subset B.$$

Notice that according to the above analysis, any *contingent* proposition satisfies the first two constraints of the above definition of *relevance*. But some

[5] In Groenendijk (1999) the notion is called 'licencing'.

contingent propositions are, of course, intuitively *irrelevant* because they are already entailed by, or inconsistent with, what is already believed by the participants of the conversation. It is only natural to expect that what is believed also influences the comparative goodness relation of answers to questions. And indeed, that turns out to be the case.

3.2 A pragmatic comparison

Although the above defined comparative notion of goodness of answers is quite appealing, it still can be the case that certain answers to a question can be better than others, although they are according to the above ordering relations predicted to be worse. It can even be the case that some responses to questions are predicted to be semantically *irrelevant*, because they do not even give a partial semantic answer to the question, but still *completely* resolve the issue. If I ask you, for instance, *What are the official languages spoken in Belgium?*, you can intuitively resolve the issue by saying *The official languages of its major neighboring countries.* The reason is, of course, that the relevance of an answer should always be determined with respect to what is believed/known by the questioner. The above answer would completely resolve my question, because I know what the major neighboring countries of Belgium are (France, Germany, and the Netherlands), and I know which official languages are spoken in those countries (French, German, and Dutch, respectively).[6]

The relevance of a question, too, depends on the relevant information state. Although the question *What is the official language of the Netherlands?* gives *semantically* rise to a non-trivial partition, I wouldn't learn much when you told me the answer. We can conclude that we should *relativize* the definitions of relevance and goodness of questions and answers to particular information states.

In comparing the 'goodness' of questions to one another, and in comparing answers, we have until now neglected what is already known or believed by the agent who asks the question. When we denote the relevant information state of the questioner by K, which is represented by a set of possible worlds, we can redefine the relevant notions. First, we can define the meaning of question $?P$ with respect to information state K, $[[?P]]_K$:

$$[[?P]]_K \quad = \quad \{\{v \in K \mid [[P]]^v \ = \ [[P]]^w \ \} \mid w \in K\}.$$

Then we can say that *question* Q is *relevant* with respect to information state K just in case Q_K is a non-singleton set. To determine whether A is a *relevant answer* to Q with respect to information state K, we first define $A_{Q,K}$, which denotes the set of cells of Q_K compatible with proposition A:

$$A_{Q,K} \quad = \quad \{q \in Q_K : \ q \cap A \neq \emptyset\}.$$

[6] Neglecting the claim of some that Frisian is an official language of the Netherlands, too.

Now we can say that A is *about* Q *with respect to* K, just in case $\bigcup A_{Q,K} = (K \cap A)$. Then we call A a *relevant answer* to Q *w.r.t.* K iff it is *contingent* with respect to K and *about* Q with respect to K. Now we are ready to compare questions and answers with respects to information states. First questions:

$$\text{Question } Q \text{ is } \textit{at least as good w.r.t. } K \text{ as } Q' \quad \text{iff} \quad Q_K \sqsubseteq Q'_K.$$

Determining the ordering relation for answers A and B that are relevant with respect to Q and K is equally straightforward:

$$A \geq_{Q,K} B \quad \text{iff} \quad (K \cap A) \subseteq (K \cap B).$$

If we want, we might also follow Groenendijk & Stokhof (1984) by also comparing answers that express the same proposition with respect to our state K. They propose that in that case one answer is better than another one, if it is *semantically* better, i.e. if it is higher on the '$>_Q$'-scale than the other one.

3.3 Limitations of qualitative comparisons

When we would relate questions and answers with respect to the relations '$\sqsubseteq_K$' and '$\geq_{Q,K}$', respectively, both relations would give rise to *partial* orderings. This is not very surprising giving our *qualitative* method used to define them. These qualitative methods are rather *coarse grained*, and this also holds for the criterium an answer should satisfy, according to the above method, to count as a *relevant* answer. Remember that according to the proposal above, answer A can only be relevant with respect to question Q and information state K if it is inconsistent with at least one cell of the partition induced by question Q_K, i.e. if $A_{Q,K} \sqsubseteq Q_K$ and $A_{Q,K} \neq Q_K$.

Although the definition of *relevance* given in the previous subsection is quite appealing, it seems that we have more intuitions about 'relevance' than this *qualitative* notion can capture. An answer can, intuitively, sometimes be relevant, although it is consistent with *all* cells of the partition. When I would ask you *Will John come?*, and you answer by saying *Most probably not*, this response counts intuitively as a very relevant answer, although it does not rule out any of the cells induced by the question. In this case the answer changes the probability distribution of the cells of the partition, but our problem also shows up when probability doesn't play a (major) role. When I ask you the *yes/no*-question *Are John and Mary sick?*, your answer *At least John is* is compatible with both answers, but still felt to be very relevant. This suggests that the notion of *relevance* of answers should be determined with respect to a more fine-grained ordering relation than our above '$\geq_{Q,K}$'.

There is also a more direct reason why the ordering relation between answers should be defined in a more fine-grained way. It is possible that one answer that is consistent with all elements of a partition can be more relevant than another (relevant) answer that is consistent with all elements of a partition, even if the one does not entail the other: The answer *(At least) John and Mary are sick*

is normally felt to be a more relevant, or informative, answer to the question *Who of John, Mary and Sue are sick?* than the answer *(At least) Sue is sick*, although less relevant than the complete answer to the question that *Only Sue is sick*. These examples suggest that we want to determine a *total* ordering relation between answers and that we should compare answers to one another in a more *quantitative* way. When probability doesn't play a role (or when all worlds have equal probability), this can simply be done by *counting* the numbers of cells of the partition the answers are compatible with, or the number of worlds compatible with what is expressed by the answers. I won't discuss such a proposal further in this paper, and turn in the next section straightaway to probabilities.

Above I have argued that we should define a more fine-grained ordering relation between answers. Something similar also holds for questions. If I want to find out who of John, Mary and Sue are sick, the question *Who of John and Mary are sick?* is felt to be more informative, or relevant, than the question *Is Sue sick?*, although none of the complete answers to the first question will solve the second issue. What this example suggests is that (i) also questions should be compared to each other with respect to a *quantitative* ordering relation, but also that (ii) to compare the usefulness of two questions with each other, we should *relate* the questions to (something like) a *third question*. Later in this paper, this third question, or problem, will show up again as a *decision problem*.

We have suggested to *extend* our *partial* ordering relations between questions and answer to *total* orderings by measuring the informativity and relevance of propositions and questions in a more *quantitative* way. But how can we do that?

4 Information and Communication Theory

4.1 The amount of information of a proposition

There turns out to be a standard way to determine the informativity of proposition that give rise to a total ordering, such that this total ordering is an extension of the ordering induced by entailment. Notice that if one proposition entails another, it is more informative to learn the former than to learn (only) the latter. That is, it would be *more surprising* to find out that the former proposition is true, than to find out that the latter is. But it doesn't seem to be a necessary condition for proposition A to be more surprising than proposition B that A entails B. All what counts, intuitively, is that the *probability* that A is true is smaller or equal than the probability that B is true. Assuming that an information state should be modeled by a probability function, P, we might say that for each proposition A, its measure of surprise can be defined as either $1 - P(A)$, or $1/P(A)$.[7] Both measures will induce the same total ordering of propositions

[7] In this paper I will assume that probabilities are assigned to worlds, and not (primarily) to propositions. Thus, a probability function, P, is a function in $[W \rightarrow [0,1]]$, such that $\sum_{w \in W} P(w) = 1$. Notice that this allows lots of worlds to have a probability of 0. A proposition, A, is represented by a set of worlds, and the probability of such a proposition, $P(A)$, is defined as $\sum_{w \in A} P(w)$.

with respect to their informativity. For reasons that will become clear later, however, we will follow Bar-Hillel & Carnap (1953),[8] and define the informativity of proposition A, $\inf(A)$, as the logarithm with base 2 of $1/P(A)$, which is the same as the negative logarithm of the probability of A:[9]

$$\inf(A) \quad = \quad \log_2 (1/P(A)) \quad = \quad -\log_2 P(A).$$

Also in terms of this notion of informativity we can totally order the propositions by means of their informativity, or measure of surprise, and it turns out that the so induced ordering corresponds exactly with the ones suggested earlier.[10]

To explain the 'inf'-notion, let us consider again the state space where the relevant issues are whether John, whether Mary, and whether Sue are sick or not. The three issues together give rise to $2^3 = 8$ relevantly different states of the world, and assuming that it is considered to be equally likely for all of them to be sick or not, and that the issues are independent of one another, it turns out that all 8 states are equally likely to be true. In that case, the informativity of proposition A equals the number of the above 3 binary issues solved by learning A. Thus, in case I learn that John is sick, one of the above three binary issues, i.e. *yes/no*-questions, is solved, and the informativity of the proposition expressed by the sentence *John is sick* $= J$, $\inf(J)$, is 1. Notice that proposition J is compatible with 4 of the 8 possible states of nature, and on our assumptions this means that the probability of J, $P(J)$, is $\frac{1}{2}$. To determine the informational value of a proposition, we looked at the negative logarithm of its probability, where this logarithmic function has a base of 2. Recalling from high-school that the logarithm with base 2 of n is simply the power to which 2 must be raised to get n, it indeed is the case that $\inf(J) = 1$, because $-\log P(J) = -\log \frac{1}{2} = 1$, due to the fact that $2^{-1} = \frac{1}{2}$. Learning that both Mary and Sue are sick however, i.e. learning proposition $M \wedge S$, has an informative value of 2, because it would resolve 2 of our binary issues given above. More formally, only in 2 of the 8 cases it holds that both women are sick, and thus we assume that the proposition expressed, $M \wedge S$, has a probability of $\frac{1}{4}$. Because $2^{-2} = \frac{1}{4}$, the amount of information learned by $M \wedge S$, $\inf(M \wedge S)$, is 2.

What if a proposition does not resolve a single one of our binary issues, like the proposition expressed by *At least one of the women is sick*, i.e. $M \vee S$? Also such propositions can be given an informative value, and in accordance with our above explanation the informative value of this proposition will be *less* than 1, because it does not resolve a single of the relevant binary issues. Notice

[8] Who in turn take over Hartley's (1928) proposal for what he calls the 'surprisal value' of a proposition.

[9] The 'inf'-value of a proposition is a function of its probability; for different probability functions, the 'inf'-value of a proposition might be different. In the text I won't mention, however, the particular probability function used.

[10] To determine this ordering it is also irrelevant what we take as the base of the logarithm. But certainly in our use of the informational value of propositions for determining the informational value of questions, the chosen base 2 will be most appealing.

that the proposition is true in 6 of the 8 states, and thus has a probability of $\frac{3}{4}$. Looking in our logarithm-table from high-school again, we can find that $-\log \frac{3}{4} = 0.415$, which is thus also the amount of information expressed by the proposition according to Bar-Hillel & Carnap's proposed measure.

In our above examples we have only looked at the special case where each of the 8 states were equally likely, and thus limited ourselves to a rather specific probability function.[11] But it should be clear that the informative value of a proposition can also be determined in case the states are not equally probable. Bar-Hillel & Carnap prove that their value function has a number of properties, and here I want to mention only the most important ones.

They note that when proposition A is already believed by the agent, i.e. when $P(A) = 1$, the amount of information gained by learning A is 0, $\inf(A) = 0$, which is a natural measure for the lower bound. The higher bound is reached when proposition A is 'learned' of which the agent believes that it cannot be true, $P(A) = 0$. In that case it holds that $\inf(A) = \infty$. The 'inf'-value of all 'contingent' propositions, i.e. of all propositions A such that $0 < P(A) < 1$, will be finite, and higher than 0.

Let us say that two propositions A and B are *independent* with respect to probability function P when $P(A \wedge B) = P(A) \times P(B)$, that is, when $P(B/A) = P(B)$. In that case it holds that $\inf(B/A) = \inf(B)$, where $\inf(B/A)$ measures the amount of information of B given that A holds, and defined as the difference between $\inf(A \wedge B)$ and $\inf(A)$:

$$\begin{aligned} \inf(B/A) &= \inf(A \wedge B) - \inf(A) \\ &= -\log_2 P(B/A). \end{aligned}$$

When A and B are independent, conjunction behaves informationally *additive*, i.e. $\inf(A \wedge B) = \inf(A) + \inf(B)$. And indeed, in our above example M and S – the propositions that Mary and Sue are sick, respectively – are independent, and both have the same 'inf'-value as J, namely 1. Thus, $\inf(M) + \inf(S) = 2$, which is exactly the 'inf'-value of $M \wedge S$, as we have observed above.

An important property of the 'inf'-function for our purposes is that it is *monotone increasing* with respect to the entailment relation. That is, if $A \subseteq B$, it holds that $\inf(A) \geq \inf(B)$. And indeed, in our above example we saw that $\inf(M \wedge S) \geq \inf(M \vee S)$. Exactly because the 'inf'-function behaves monotone increasing with respect to the entailment relation, the total ordering relation induced by the 'inf'-function has the nice property that it is an *extension* of the partial ordering relation induced by the entailment relation. The entailment

[11] The kind of probability function we used is closely related to Carnap's (1950) *objective* probability function, and also used in Bar-Hillel & Carnap (1953), to define an *objective* notion of amount of the *semantic* information of a proposition. But the way they define the informativity of a proposition does obviously not demand the use of such an objective probability function. The informative value of a proposition is always calculated with respect to a particular probability function, and this probability function might well be *subjective* in the sense that it represents the beliefs of a particular agent.

relation and the ordering relation induced by the 'inf'-function are even closer related to each other: if with respect to every probability function it holds that $\inf(A) \geq \inf(B)$, then it will be the case that A semantically entails B. What this suggests is that the *semantic entailment* relation is an *abstraction* from the more *pragmatically oriented amount-of-information relation.*[12]

4.2 The entropy of a question

Now that we have extended the ordering relation between propositions with respect to their information values to a total relation, the question arises whether something similar can be done for *questions*. As before, I will think of questions as semantic object, and in particular as partitions of the state space.

To determine the informative value of a question, we will again follow the lead of Bar-Hillel & Carnap (1953). They discuss the problem how to determine the *estimated* amount of information conveyed by the outcome of an *experiment* to be made. They equate the value of an experiment with its estimated amount of information, and they assume that the possible outcomes denote propositions such that the set of outcomes are mutually exclusive and jointly exhaust the whole state space. In other words, they assume that the set of possible outcomes *partitions* the set of relevant states. This suggests, obviously, that we can also equate the informative value of a *question* with the estimated amount of information conveyed by its (complete) answers. The estimated amount of information of the answers will simply be the *average* amount of information of the answers. For reasons that will become clear soon, I will denote the informative value of question Q by $E(Q)$, which will be defined as follows:

$$E(Q) \quad = \quad \sum_{q \in Q} P(q) \times \inf(q).$$

To strengthen our intuitions, let us look again at the case where we have 8 relevantly different states of the world, such that each of the states are equally likely to be true. Consider now the question *Who of John, Mary and Sue are sick?*

[12] Of course, the semantic entailment relation (a partial ordering) is defined in terms of *meaning*, while the total ordering relation is defined in terms of a different kind of concept. Some early proponents of communication theory, however, didn't make a great effort to keep the concepts separate. Norbert Wiener (1950), for instance, takes *amounts of information* and *amount of meaning* to be equivalent. He says, "The amount of meaning can be measured. It turns out that the less probable a message is, the more meaning it carries, which is entirely reasonable from the standpoint of common sense." But, to quote Dretske (1999, p. 42) "It takes only a moment's reflection to realize that this is *not* 'entirely reasonable' from the standpoint of common sense. There is no simple equation between meaning (or amount of meaning) and information (or amount of information) as the latter is understood in the mathematical theory of information. The utterance *There is a gnu in my backyard* does not have *more meaning* than *There is a dog in my backyard* because the former is, statistically, less probable."

Notice that any complete answer to this question will reduce our 8 possibilities to 1. Thus, any complete answer, q_i, will have an 'inf'-value of 3, i.e. it will resolve all three of the relevant binary issues. But if each answer to the question has an informative value of 3, the *average* amount of information conveyed by the answers, and thus the informative value of the question, $E(Q)$, should also be 3. And indeed, because each of the complete answers has a probability of $\frac{1}{8}$ to be true, the informative value of the question is according to the above formula equated with $(\frac{1}{8} \times 3) + ... + (\frac{1}{8} \times 3) = 8 \times (\frac{1}{8} \times 3) = 3$. In general it will hold that when we have n mutually exclusive answers to a question, and all the answers are considered to be equally likely true, the informative value of the question can simply be equated with the informative value of each of its answers, which is $-\log_2 \frac{1}{n} = \log_2 n$. The informative value of the question *Will the outcome of the flipping of an unbiased coin be heads?*, for instance, will be 1, because the question has 2 answers, which are by assumption equally likely to be true.

What if not all of the n answers are equally likely to be true? In that case some answers have a higher informative value than $\log_2 n$, and others have a lower one. It turns out, however, that the *average* amount of information conveyed by the answers will in that case be *lower* than in case the answers are equally likely to be true. Consider for instance the flipping of a *biased coin*, whose chance to come up heads after flipping is $\frac{3}{4}$. Because the 'inf'-value of outcome/answer *Heads* is in that case $-\log_2 \frac{3}{4} = 0.415$, and the 'inf'-value of answer *Tails* is $-\log_2 \frac{1}{4} = 2$, the *average* amount of information of the answers is $(\frac{3}{4} \times 0.415) + (\frac{1}{4} \times 2) = 0.811 < 1$. Thus, although one of the answers has an informative value that is 2 times as high as the informative values of the outcomes/answers in case of an unbiased coin, the average amount of information of the answers turns out to be lower.

This is in general the case; the informative value of question Q defined as above is *maximal* just in case the answers are all equally likely to be true. And this seems to confirm to our intuitions. If you want to be sure to find out after 3 *yes/no*-questions which of the 8 states of our toy-example actually obtains, we should ask the three *yes/no*-questions which have maximal E-value. That is, we should ask for each individual separately whether he or she is sick, which all have an 'inf'-value of 1, and we should not ask *risky* questions that might, but need not, convey more information, like *Are John and Mary the ones who are sick?* In fact, we might even define the *risk* of question Q which has n different possible answers, as (a function of) the difference between the E-value of the n-ary question with maximal informative value, i.e. with an E-value of $\log_2 n$, and $E(Q)$.

Having defined when a question has its maximal informative value, we now would like to know under which circumstances it reaches its *minimal* value. Intuitively, a question is (at least) valueless in case you already know the answer to the question. And, unsurprisingly, this is what comes out; $E(Q) = 0$ just in case only one answer has a positive probability (and thus has the probability 1), and for all other cases the question has a value strictly higher than 0.

Our aim was to define a value of questions (partitions) that allows us to extend the partial ordering on questions induced by the '$\sqsubseteq$' relation to a total

ordering. We have succeeded in doing that: it always will be the case that when $Q \sqsubseteq Q'$, it will also be the case that $E(Q) \geq E(Q')$. Moreover, as a special case of a theorem stated in section 5 it will be the case that if $E_P(Q) \geq E_P(Q')$ with respect to all probability functions P, it holds that $Q \sqsubseteq Q'$.

We have defined the informative value of questions in the same way as Bar-Hillel & Carnap (1953) defined the value of doing an experiment. As they have noted themselves, the way this value is defined is *formally* exactly analogous to the way the *entropy* of a *source*, i.e. coding system, is defined by Shannon (1948) in his *Communication Theory*. This is why we denoted the informative value of question Q by $E(Q)$, and from now on I will call the informative value of a question simply its entropy. In Communication Theory 'entropy' is the central notion, because engineers are mostly interested in the issue how to device a coding system such that it can transmit *on average* as much as possible information via a particular *channel*. Although we have defined the entropy of a question formally in the same way as Shannon defined the entropy of a source, there is an important differences between Shannon's original use of the formalism within Communication Theory on the one hand, and Bar-Hillel & Carnap's and our application of it on the other: Shannon looked at things from a purely *syntactic* point of view while we interpret notions like 'informativity' and 'entropy' in a *semantic/pragmatic* sense.

4.3 Conditional entropy, and the informative value of expressions

Although we have followed Bar-Hillel & Carnap in making a different use of the formalism Shannon invented than originally intended, this doesn't mean that we are not allowed to 'borrow' some mathematical results Shannon proved for his theory of entropy. In particular, we can make use of what in Communication Theory is known as *conditional entropy*, and of what is sometimes called *Shannon's inequality*, to determine the estimated reduction of uncertainty due to getting an answer to a question.

To use those notions, we first have to say what the *joint entropy* is of two questions, Q and Q', $E(Q, Q')$, is, where both Q and Q' are as usual represented by partitions:

$$E(Q, Q') \quad = \quad \sum_{q \in Q} \sum_{q' \in Q'} P(q \cap q') \times \mathrm{Inf}(q \cap q').$$

It should be clear that this joint entropy of Q and Q' is equivalent to the entropy of $Q \sqcap Q'$, $E(Q \sqcap Q')$, where $Q \sqcap Q' \stackrel{def}{=} \{q \cap q' : q \in Q \ \& \ q' \in Q' \ \& \ q \cap q' \neq \emptyset\}$.

Until now we have defined the entropy of a question with respect to a set of ways the world might be. Notice that the set of worlds consistent with what is believed, $\{w \in W : P(w) > 0\}$, corresponds itself also to a partition, namely the most fine-grained partition $\{\{w\} : P(w) > 0\}$. Calling this latter partition B, also this partition can be thought of as a question that has a certain entropy, $E(B)$.

Let us now assume that the agent learns answer q to question Q. What is then the entropy of B *conditional* on learning q, $E_q(B)$? The definition of this conditional entropy can be easily given:

$$E_q(B) \quad = \quad \sum_{b \in B} P(b/q) \times \inf\,(b/q),$$

and measures the entropy of, or uncertainty in, B when it is known that the answer to Q is q_i. In terms of this notion we might now define the *entropy of B conditional on Q*, $E_Q(B)$. This is defined as the *average* entropy of B conditional on learning an answer to question Q:

$$\begin{array}{rcl} E_Q(B) & = & \sum_{q \in Q} P(q) \times E_q(B) \\ & = & \sum_{q \in Q} P(q) \times \sum_{b \in B} P(b/q) \times \inf\,(b/q) \\ & = & \sum_{q \in Q} \sum_{b \in B} P(q \wedge b) \times \inf\,(b/q). \end{array}$$

Now it can be shown that for any two partitions X and Y of the same set of worlds, it holds that $E(X,Y) - E(X) = E_X(Y)$:

$$\begin{array}{rcl} E(X,Y) - E(X) & = & -\sum_{x \in X} \sum_{y \in Y} P(x \wedge y) \times logP(x \wedge y) + \\ & & \sum_{x \in X} P(x) \times logP(x) \\ & = & \sum_{x \in X} \sum_{y \in Y} P(x \wedge y) \times logP(x) - \\ & & \sum_{x \in X} \sum_{y \in Y} P(x \wedge y) \times logP(x \wedge y) \\ & = & \sum_{x \in X} \sum_{y \in Y} P(x \wedge y) \times [logP(x) - logP(x \wedge y)] \\ & = & \sum_{x \in X} \sum_{y \in Y} P(x \wedge y) \times log \frac{P(x)}{P(x \wedge y)} \\ & = & \sum_{x \in X} \sum_{y \in Y} P(x \wedge y) \times inf(y/x) \\ & = & E_X(Y). \end{array}$$

A similar calculation shows that $E(X,Y) - E(Y) = E_Y(X)$, and thus that $E_X(Y) + E(X) = E_Y(X) + E(Y)$. Notice that thus in particular it holds for our two partitions Q and B that $E_Q(B) = E(Q,B) - E(Q)$. I will just state, and not show, *Shannon's inequality*, which says that for any two partitions X and Y of the same state space, it holds that

$$E_X(Y) \quad \leq \quad E(Y),$$

where the two values are the same exactly when the two issues are completely *orthogonal* to one another, i.e. when the issues are *independent*. Notice that this means that the entropy of $Q \sqcap Q'$ only equals the entropy of Q plus the entropy of Q' in case the partitions are fully independent. That the entropy of the combined question is only in these special cases equal to the sum of the entropies of the questions separately, confirms to our intuition that *on average* we learn less by getting an answer to the combined question *Who of John and Mary will come to the party?*, than by getting two separate answers to both questions *Will John come to the party?* and *Will Mary come to the party?*, when John only, but *not* if and only, comes when Mary comes.

Shannon's inequality will turn out to be a nice property of what I will call the average information gained from the answer to a question. To define this notion, let us first define what might be called the *Informational Value* of answer q, with respect to partition B, $IV_B(q)$, as the reduction of entropy, or uncertainty, of B when q is learned:[13]

$$IV_B(q) \quad = \quad E(B) - E_q(B).$$

Because learning q might flatten the distribution of the probabilities of the elements of B, it should be clear that $IV_B(q)$ might have a negative value. Still, due to Shannon's inequality, we might reasonably define the informational value of question Q, the *Expected Informational Value* with respect to partition B, $EIV_B(Q)$, as the *average* reduction of entropy of B when an answer to Q is learned:

$$\begin{array}{rcl} EIV_B(Q) & = & \sum_{q \in Q} P(q) \times IV_B(q) \\ & = & \sum_{q \in Q} P(q) \times [E(B) - E_q(B)] \\ & = & E(B) - [\sum_{q \in Q} P(q) \times E_q(B)]^{14} \\ & = & E(B) - E_Q(B) \end{array}$$

The difference between $E(B)$ and $E_Q(B)$ is also known as the *mutual information* between B and Q, $I(B, Q)$. Shannon's inequality tells us now that our *average* uncertainty about B can never be increased by asking a question, and it remains the same just in case Q and B are orthogonal to each other. In the latter case we might call the question *irrelevant*.

To strengthen our intuitions, let us look at our toy-example again. Recall that 8 worlds were at stake, and all the 8 worlds had the same probability. In that case, learning which element of B obtains, i.e. learning what the actually world is, gives us 3 bits of information, and thus $E(B) = 3$. Remember also that learning the answer to the *yes/no*-question *Is John sick?* will give us 1 bit of information, i.e. $E(\text{Sick(j)?}) = E(Q) = 1$, because each answer to the question is equally likely true, and from both answers we would gain 1 bit of information. It's almost equally easy to see that for both answers to the question, the entropy of B *conditional* on learning this answer q, $E_q(B)$, is also 1, and thus that the *average* reduction of uncertainty due to an answer to Q, $E_B(Q)$ is 1, too. It follows that thus the expected information value, $EIV_B(Q)$, is $E(B) - E_Q(B) = 3 - 1 = 2$. The same result is achieved when we determine $EIV_B(Q)$ by taking the average difference between $E(B)$ and $E_q(B)$ for both answers q, because both answers are equally likely, and for both it holds that $E(B) - E_q(B) = IV_B(q) = 2$.

We have defined the expected informational value of question Q with respect to partition B, $EIV_B(Q)$, as the average reduction of entropy of B when an answer to Q is given, i.e. as the difference between $E(B)$ and the *conditional*

[13] A similar notion was used by Lindley (1956) to measure the informational value of a particular result of an experiment.

[14] This step is allowed because the *unconditional* entropy of B, $E(B)$, does not depend on any particular element of Q.

entropy $E_Q(B)$. And to make sure that this is always positive, we have made use of Shannon's inequality. But notice that the entropy of B conditional on Q, $E_Q(B)$, is simply the same as the entropy of Q, $E(Q)$, itself. But this means that the expected informational value of Q with respect to B, $EIV_B(Q)$, can also be defined as the difference between the entropy of B and the entropy of Q, $E(B)-E(Q)$. Notice also that we don't have to make use of Shannon's inequality to see that for any question Q it holds that $EIV_B(Q)$ will never be negative. The reason is that for any question Q it holds that $B \sqsubseteq Q$, and we have noted already that in that case it will hold that the entropy of B will be at least as high as the entropy of Q: $E(B) \geq E(Q)$. But if we can assure that the informational value of a question is non-negative without making use of Shannon's inequality, why did we define the value of a question in such a roundabout way via the conditional entropy of B given Q?

4.4 Deciding between Hypotheses

The reason is that we don't want to restrict ourselves to the *special case* where in the end we want to have *total* information about the world, where we have completely reduced all our uncertainty. Remember that partition B was the most fine-grained partition possible; the elements of B were *singleton sets* of worlds. Because the entropy of Q measures the average uncertainty about how the world looks like when we've got an answer to Q, this measure, $E(Q)$, is only the same as the entropy of B conditional on Q, $E_Q(B)$, because the elements of our special partition B correspond one-to-one to the worlds.[15]

But now suppose that we need not to know how *exactly* the world looks like, but rather just want to find out which of the mutually exclusive and exhaustive set of *hypotheses* in the set $H = \{h_1, ..., h_n\}$ is true, where the h_i's denote arbitrary propositions. The problem now is to determine the value of question Q with respect to this other partition H, $EIV_H(Q)$, and this is in general not equal to $E(H) - E(Q)$. To determine the value $EIV_H(Q)$, we need to make use of the conditional entropy of H given (an answer to) Q.

Notice that Shannon's inequality tells us now also something informative; $EIV_H(Q)$ will never be negative, although it need not be the case that $H \sqsubseteq Q$. And not only for our special partition B, but also for any other issue H we can determine when question Q is informationally relevant. Question Q is *informationally relevant* with respect to a set of hypotheses H just in case the true answer to Q is expected to reduce the uncertainty about what the true hypothesis of H is, i.e. $EIV_H(Q) > 0$.

This notion of 'informational relevance' is important when an agent is fronted with the *decision problem* which of the mutually exclusive hypotheses $\{h_1, ..., h_n\}$ he should choose. In case the agent only cares about the issue which of the hypotheses is true, and that all ways of choosing falsely are equally bad, the risk of choosing depends only on the uncertainty about what the right hypothesis

[15] More in general, it holds that for two partitions Q and Q', if $Q \sqsubseteq Q'$, then $E_Q(Q') = E(Q')$.

is. It seems natural to advice him in these circumstances always to choose that hypothesis that has the highest prior probability. But this means that the *risk* of choosing depends entirely on the entropy of H, $E(H)$. And indeed, the flatter the distribution of the probabilities of the hypotheses is, the more risky the choice will be.

Notice that asking a question, and thereby expecting to get an answer (that is true), might reduce the entropy of H, i.e. the uncertainty about which hypothesis is true, and thus also the *risk* of the decision, even if all answers to the question are compatible with all hypotheses. But this means that even if no answer to the question will eliminate a single hypothesis, it might still be *useful*, or *relevant*, to ask the question. Indeed, at this point it seems only natural to equate the *usefulness* of question Q with respect to the *decision problem* which of the hypotheses of H should be chosen, with the reduction of uncertainty about H due to Q, i.e. $EIV_H(Q)$. Moreover, we can say that question Q is *relevant* with respect to H just in case $EIV_H(Q)$ is strictly higher than 0.

Thus, instead of the *partial* order between questions induced by the relation '$\sqsubseteq$', we can now determine a *total* order. We say that if $Q \neq Q'$, question Q is *better* than question Q' with respect to hypotheses H, $Q >_H Q'$, just in case the expected information value of Q is higher than the value of Q', or, if both are the same, the former is less fine-grained than the latter:[16]

$$Q >_H Q' \quad \text{iff} \quad \begin{array}{ll} (i) & EIV_H(Q) > EIV_H(Q'), \text{ or} \\ (ii) & EIV_H(Q) = EIV_H(Q') \text{ and } Q \sqsupseteq Q'. \end{array}$$

Just as the usefulness, and relevance, of question Q with respect to decision problem H can be defined in terms of $EIV_H(Q)$, we can also define the usefulness, and relevance, of *assertion* A with respect to decision problem H in terms of the information values of answers. That is, we can propose to equate the usefulness of assertion A with respect to issue H with $IV_H(A)$, and we can say that assertion A is *relevant* just in case $IV_H(A) > 0$. Notice that the thus defined notion of relevance predicts that many assertions are relevant, although they are (falsely) predicted to be irrelevant according to the *qualitative* notion of relevance used above. Moreover, our newly defined notion of relevance still has the nice property that it can explain why an assertion is felt to be *irrelevant* although it still is *informative*. For instance, if the issue is who of John and Mary are sick, and we look at our toy-example again where the sickness of John, Mary and Sue are independent of each other, the assertion *Sue is sick* is rightly predicted to be irrelevant, although it does eliminate some possible worlds.

Now we can also turn our *partial* order between answers induced by the relation '$>_Q$', to a *total* order (although it is not an extension of it). We say that assertion A is *better* than assertion B with respect to hypotheses H, $A >_H B$, just

[16] As before, I assume always a particular probability function. If we don't do that, the following general fact can be proved: Denote the expected utility value of Q with respect to H and probability function P: If $Q \sqsubseteq Q'$, then for all P : $EIV_H^P(Q) \geq EIV_H^P(Q')$. We will see in section 5 what is needed to strengthen this fact to the stronger *if and only if* statement.

in case the informational value of A, $IV_H(A)$, is higher than the corresponding value of B, $IV_H(B)$, or, in case both are the same, the former should be less surprising than the latter:

$$A >_H B \quad \text{iff} \quad \begin{array}{ll} (i) & IV_H(A) > IV_H(B), \text{ or} \\ (ii) & IV_H(A) = IV_H(B) \text{ and } inf(A) < inf(B). \end{array}$$

Thus, if A reduces the entropy of H more than B does, it is a better answer to 'question' H than B.

Notice that according to our definition of the relevance of an assertion, an assertion is predicted to be irrelevant when it flattens the probability distribution of the hypotheses. In such cases the assertion indeed has the effect that it doesn't make the decision any easier. Intuitively, however, this doesn't mean that thus the assertion is felt to be irrelevant. The assertion seems to *be* relevant exactly *because* it makes the decision more risky. This wrong prediction can, fortunately, be removed easily. Just say that A is *relevant* with respect to H exactly when the acceptance of A *changes* the probability distribution of the hypotheses, i.e. when $IV_H(A) \neq 0$.

4.5 Limitations of the analysis in terms of entropy

The measure of usefulness and relevance of questions and assertions with respect to a decision problem that we have defined above is, I think, reasonable for some, but also *only* reasonable for some kinds of decision problems. First, in our description of decision problems, we only looked at problems where the choice between a set of *hypotheses* is at stake. We would like to extend the analysis from the choice between hypotheses, to choices between more general kinds of *actions*. Extending our analysis from decisions between hypotheses to decisions between actions need not yet worry us. It doesn't seem to be completely unreasonable to represent actions as propositions; an action is true in a world just in case the result of the action is true in that world. Indeed, in the well respected decision theory of Jeffrey (1965), actions are represented by propositions.

What is more problematic for the way we have analyzed the usefulness of questions and answers in this section is that once we think of a decision problem as consisting of a set of actions, it seems only natural to assume that the decision depends not only on the *probabilities* involved, but also on the *desirabilities*, or *utilities*, of the states that result when the various actions would be chosen. But once desirabilities enter the picture, it is obvious that our analysis of the usefulness of questions and answers can no longer be defined simply in terms of the dependencies between certain probability distributions, i.e. in terms of conditional entropies.

Consider, for instance, the decision problem faced by airpilot Smith who wonders whether he should drop the bomb, with the reasonable chance to trigger a world-war, or not dropping the bomb, and thereby missing an excellent chance to strike a potential future enemy in war, and getting a scolding for this by

his commanding officer. It is clear that Smith's desirabilities of the expected outcomes of the relevant actions will heavily influence his decision.

Even if the relevant actions just involves a choice between a set of hypotheses, the most probable hypothesis is not always the one that intuitively is preferable. The reason is that choosing this hypothesis might give rise to very nasty consequences. Consider, for instance, scientist Jones who is facing the dilemma between choosing the generally accepted theory h_1 and working in this framework, or choosing the alternative theory h_2 that he thinks is somewhat more likely to be true, but that has a very bad reputation among his fellow researchers. Because Jones knows that choosing h_2 will turn him into a black sheep of his family whose papers will never be read, even the more purists among us could understand Jones' choice for theory h_1.

Let me give a simple example showing that the reduction of entropy of the relevant set of hypotheses/actions does not always measure the usefulness of questions and assertions in a satisfying way. Consider John, who wonders whether he should go to the party tonight, or not. His decision depends almost entirely on whether Mary will go, because he is secretly in love with Mary, and believes that going to the party is his only chance to meet her. He prefers meeting her tonight, to not meeting her, but if Mary won't go, he prefers to stay home. But going to the party when Mary comes too obviously involves a risk; perhaps Mary will turn him down when he makes his advances. We might say that in this situation 4 different states (worlds) are involved: one world, w_1, where Mary goes to the party, John will go, too, he will try his luck, and is successful; a world, w_2, where Mary goes, John goes, he tries his luck, and is unsuccessful; world w_3, where Mary won't go to the party, and thus neither does John, but where the counterfactual statement holds that when John *would* try his luck, he *would* be successful, and w_4 which is similar to w_3 except that in this world the counterfactual would be false. On the additional assumption that John thinks all worlds are equally likely to come out true, that he doesn't care about what Mary would do if they don't go to the party, and that John has a negative attitude towards taking risks, we might represent his decision problem by the following table:

World	Probability	Desirability
w_1	1/4	12
w_2	1/4	2
w_3	1/4	8
w_4	1/4	8

In this case, it is *relevant*, intuitively, for John to learn that the above mentioned counterfactual statement is true. It is, however, easily seen that learning the proposition expressed by this statement, $\{w_1, w_3\} = A$, does *not* change the entropy of the decision problem that can be represented by $\{\{w_1, w_2\}, \{w_3, w_4\}\} = H$. That is, $IV_H(A) = E(H) - E_A(H) = 0$, because learning A does not change the probability distribution of the elements of H, i.e. both $E(H)$ and $E_A(H)$ have a value of 1.

In a similar way, it also seems relevant for John to know the answer to the question whether he would be successful if he tried, that is, to learn which element of the partition $\{\{w_1, w_3\}, \{w_2, w_4\}\}$ is true. It is straightforward to check, however, that not only the positive answer to the question, A, but also the negative answer, $\neg A$, has no effect on the probability distribution of the elements of H. Representing the question whether the counterfactual is true or not by Q, it is thus predicted that also $EIV_H(Q) = E(H) - E_Q(H) = 0$. We can conclude that the value $EIV_H(Q)$ is at least not always the proper measure to determine the relevance of a question with respect to a decision problem.

What we need, or so it seems, is a measure that not only looks at the probabilities, but also at the desirabilities involved. In the next section we will define such a measure by looking seriously at statistical decision theory.

5 Utility Values of Questions and Answers

5.1 Utilies of answers and expected utilities of questions

In Savage's (1954) decision theory, actions are taken to be primitives, and if we assume that the utility of performing action a in world w is $U(a, w)$, we can say that the *expected utility* of action a, $EU(a)$, with respect to probability function P is

$$EU(a) \quad = \quad \sum_w P(w) \times U(a, w).$$

Let us now assume that our agent, John, faces a *decision problem*, i.e. he wonders which of the alternative actions in $\mathcal{A}$ she should choose. A decision problem of an agent can be modeled as a triple, $\langle P, U, \mathcal{A}\rangle$, containing (i) the agent's probability function, P, (ii) his utility function, U, and (iii) the alternative actions he considers, $\mathcal{A}$. You might wonder why we call this a decision *problem*; shouldn't the agent simply choose the action with the highest expected utility? Yes, he should, if he *chooses now*. But now suppose that John doesn't have to choose now, but that he has the opportunity to first receive some useful information by *asking question* Q.

Before we can determine the utility of Q, we first have to say how to determine the expected utility of an action conditional on learning some new information. For each action $a \in \mathcal{A}$, its conditional expected utility with respect to new proposition C, $EU(a, C)$ is

$$EU(a, C) \quad = \quad \sum_w P(w/C) \times U(a, w).$$

When John learns proposition C, he will of course choose that action in $\mathcal{A}$ which maximizes the above value. Then we can say that the utility value of making an informed decision conditional on learning C, UV(Learn C, choose later), is the expected utility conditional on C of the action that has highest expected utility:

$$UV(\text{Learn } C\text{, choose later}) \quad = \quad max_{a \in \mathcal{A}} EU(a, C).$$

In terms of this notion we can determine the value, or *relevance*, of the assertion C. Referring to a^* as the action that has the highest expected utility according to the original decision problem, $\langle P, U, \mathcal{A} \rangle$, i.e. $max_{a \in \mathcal{A}} EU(a) = EU(a^*)$, we can determine the *utility value* of the *assertion* C, $UV(C)$, as follows:

$$UV(C) \quad = \quad max_{a \in \mathcal{A}} EU(a, C) - EU(a^*, C).$$

This value, which in statistical decision theory (cf. Raiffa & Schlaifer, 1961) is known as the *value of sample information* C, $VSI(C)$, can obviously never be negative. In fact, it predicts that an assertion only has a positive utility value in case it influences the action that John will perform. And indeed, it seems natural to say that a cooperative participant of the dialogue only makes a *relevant* assertion in case it makes John *change* his mind with respect to which action he should take. It also seems not unreasonable to claim that in a cooperative dialogue one assertion, A, is 'better' than another, B, just in case the utility value of the former is higher than the utility value of the latter, $UV(A) > UV(B)$.

In terms of the utility value of assertions/answers, we can now determine the utility values of *questions*. Suppose that question Q is represented by the partition $\{q_1, ..., q_n\}$. Just like in section 4 we defined the informative value, or entropy, of a question as the *expected*, or *average*, informative value of its answers, in this case we can determine the *expected* utility value of a question, $EUV(Q)$ as the *average* utility value of the possible answers:

$$EUV(Q) \quad = \quad \sum_{q \in Q} P(q) \times UV(q).$$

Notice that this value, which in statistical decision theory is known as the *expected value of sample information*, $EVSI$, will never be negative. In fact, the value will only be 0 in case no answer to the question would have the result that the agent will change his mind about which action to perform, i.e. for each answer $q \in Q$ it will be the case that $max_{a \in \mathcal{A}} EU(q, a) = EU(q, a^*)$. In these circumstances the question really seems irrelevant, and it thus seems natural to say that question Q is *relevant* just in case $EUV(Q) > 0$. It should be obvious that this measure function also totally orders all questions with respect to their expected utility value.

It is of some interest to see that we can determine the expected utility value of questions also in another way. According to this alternative way of determining the value of questions, we first have to determine the utility value of *choosing now*. The utility value of choosing now is defined as the expected utility of the action which has the highest expected utility according to the original decision problem, i.e. with respect to the original probability function:

$$UV(\text{Choose now}) \quad = \quad max_{a \in \mathcal{A}} EU(a).$$

Now we can determine the expected utility value of choosing after you learn the answer, EUV(Learn answer, choose later), in terms of UV(Learn q, choose later), by averaging over the answers to the question:

$$\begin{aligned} EUV(\text{Learn answer, ch. later}) &= \textstyle\sum_{q \in Q} P(q) \times UV(\text{Learn } q\text{, ch. later}) \\ &= \textstyle\sum_{q \in Q} P(q) \times max_{a \in \mathcal{A}} EU(a, q). \end{aligned}$$

The expected utility value of question Q, $EUV^{\dagger}(Q)$, is now defined as the difference between the expected utility value of choosing after you got the answer, and the utility value of choosing now:

$$EUV^{\dagger}(Q) \quad = \quad EUV(\text{Learn answer, choose later}) - UV(\text{Choose now}).$$

It can be easily shown that the second way of determining the expected utility value of a question gives rise to the same result as determining the expected utility value of a question according to the first way, i.e. $EUV^{\dagger}(Q) = EUV(Q)$:[17]

$$\begin{aligned} EUV^{\dagger}(Q) &= EUV(\text{Learn answer, choose later}) - UV(\text{Choose now}) \\ &= [\textstyle\sum_{q \in Q} P(q) \times UV(\text{Learn } q\text{, choose later})] - UV(\text{Choose now}) \\ &= [\textstyle\sum_{q \in Q} P(q) \times max_{a \in \mathcal{A}} EU(a, q)] - EU(a^*) \\ &= [\textstyle\sum_{q \in Q} P(q) \times max_{a \in \mathcal{A}} EU(a, q)] - [\textstyle\sum_{q \in Q} P(q) \times EU(a^*, q)] \\ &= \textstyle\sum_{q \in Q} P(q) \times [max_{a \in \mathcal{A}} EU(a, q) - EU(a^*, q)] \\ &= \textstyle\sum_{q \in Q} P(q) \times UV(q) \\ &= EUV(Q). \end{aligned}$$

According to the qualitative comparison method of section 3, one question, Q, is better than another question, Q', just in case the former *entails* the latter, that is, in case the partition Q is *finer* than the partition Q': $\forall q \in Q : \ \exists q' \in Q' : q \subseteq q'$. We have seen in section 4 that measuring the expected informational value of questions, $EIV_H(Q)$, in terms of reduction of entropy of the set of hypotheses H, accords with the qualitative measurement, in the sense that when Q is a finer partition than Q', it also holds that Q will have a greater expected informational value, $EIV_Q(H) \geq EIV_{Q'}(H)$, whatever the set of hypotheses is. Now we can ask a similar question with respect to the question's expected utility value. Denoting by $EUV_{DP}(Q)$ the expected utility value of Q with respect to decision problem DP, Marschak & Radner (1972) have proved as a special case of Blackwell's (1953) theorem the following strong, but also very appealing theorem:

Theorem $\qquad Q \sqsubseteq Q' \quad \text{iff} \quad \forall DP : \ EUV_{DP}(Q) \geq EUV_{DP}(Q').$

The 'only if' part is natural, and shows that it is never irrational (if collecting evidence is cost free) trying to get more information to solve one's decision problem. This part was already implicitly assumed by Savage (1954) and Raiffa & Schlaifer (1961), and was explicitly proved by Good (1966) to follow from the Bayesian principle of maximizing expected utility.[18]

[17] Where a^* is again the action which maximizes expected utility in the original decision problem.

[18] But see Skyrms (1990), who traces this result back all the way to an unpublished manuscript of Ramsey.

The 'if' part is more surprising, and it suggests that the *semantic* entailment relation between questions is an *abstraction* from the more *pragmatic* usefulness relation of questions. The proof is based on the idea that when two partitions are qualitatively incomparable, one can always find a pair of decision problems such that the first partition has a higher expected utility value than the second one according to one decision problem, and a lower expected utility value than the second one according to the other decision problem.

Given this result for questions, one might expect that something similar holds for *assertions*. We have seen in section 4.1 that whenever $A \subseteq B$, it also holds that $\inf(A) \geq \inf(B)$. In section 4.3, however, we saw that in such circumstances it still might be that $IV_H(B) > IV_H(A)$, i.e. the informational value of a proposition does not behave monotone increasingly with respect to the (ordering induced by the) classical entailment relation between propositions. Still, it might be the case that stronger propositions always *do* have a higher utility value. But in fact, they do not. The utility value of choosing now, UV(Choose now), might be higher than the utility value of first learning proposition C, and then choosing later, UV(Learn C, choose later), because from learning C I might learn that my worst nightmare has come out true, and that I have to perform an action that I otherwise never would have performed.

If neither the informative value of proposition A, $IV_H(A)$, nor its utility value, $UV(A)$, behaves monotone increasing with respect to the '$\subseteq$'-relation, perhaps they *do* behave monotone increasing with respect to one another. But also that is in general not the case, as it should be according to our argumentation in section 4.5.

First, it might be the case that learning a proposition that doesn't change the entropy, still effects a change of mind. Look at the following matrix for the example discussed in section 4.5, but now for a Savage-style decision theory:

	World	Prob	John goes	doesn't go
Mary comes	w_1	1/4	12	0
Mary comes	w_2	1/4	2	0
Mary doesn't come	w_3	1/4	0	8
Mary doesn't come	w_4	1/4	0	8

Looking at the matrix, we can equate the action *John goes* with the worlds where this action has a higher utility than the alternative action. Thus, the action corresponds in this case with the proposition $\{w_1, w_2\}$. The decision problem which action John should perform can thus be represented by the partition $\{\{w_1, w_2\}, \{w_3, w_4\}\} = H$. Note that due to the fact that all worlds have an equal probability, the informational value of proposition $\{w_1, w_3\} = A$ is 0, $IV_H(A) = 0$. Still, learning the proposition has a positive utility value, i.e. $UV(A) > 0$, because learning the proposition would have the result that John changes his mind. Facing his original decision problem, John would decide not to go to the party, because that action has the highest expected utility, UV(Choose now) $= max_i EU(a_i) = EU$(doesn't go) $= 4$. When he would learn proposition $A = \{w_1, w_3\}$, however, John would change his mind, because EU(John goes, A) $=$

$6 > EU(\text{doesn't go}, A) = 4$. Due to this latter inequality, together with the fact that the action *doesn't go* is the one that would originally have been chosen, it follows that also $UV_H(A) = 2 > 0$. This shows that information can be useful with respect to a decision problem, although it doesn't reduce the problem's entropy.

With the help of the same matrix we can also show that a proposition might reduce the entropy of a decision problem, although it doesn't have a positive utility value. We just have to find a proposition that strengthens the choice for the action/hypothesis that would have been chosen anyway, in our case for action/hypothesis $\{w_3, w_4\}$. Of course, any subset of this action/hypothesis will do this trick.

5.2 Decision between hypotheses

In section 4 our problem was to choose an hypothesis from set H, and base this decision only on the probabilities involved. A decision problem can in such cases be modeled by a pair like $\langle P, H\rangle$. As for all kinds of decision problems, we are interested in two kinds of questions: (i) What is the hypothesis the agent should go for? and (ii) What kind of question should the agent ask to make a better informed decision concerning the hypotheses? The answer to the first question seems obvious; the hypothesis the agent should choose is the hypothesis which is most likely to be true, i.e. the hypothesis with the greatest probability. The second question is somewhat more difficult to answer. Let me now show, following Marschak (1974a), that this is a special case where the decision problem should be modeled by a triple like $\langle P, U, \mathcal{A}\rangle$, as in the previous section.

We have assumed in the previous section that a decision problem partly consists of a set of alternative actions, and that each action $a \in \mathcal{A}$ has a utility in a world w, $U(a, w)$. Let us now assume that the set of alternative actions, $\mathcal{A}$, is such that for each world w there is always exactly one action $a \in \mathcal{A}$ such that $\forall a' \in (\mathcal{A} - \{a\}) : \ U(a, w) > U(a', w)$. This means that the set of alternative actions *partitions* the set of worlds; to each action $a \in \mathcal{A}$ there corresponds a cell of the partition, and in each world of this cell a is the unique best action to do. This set corresponds of course exactly to a set of mutually exclusive and jointly exhaustive hypotheses, $H = \mathcal{A}^*$, that we used in section 4 to measure the informational values of questions and answers when we define this partition as follows:

$$H = \mathcal{A}^* = \quad \{\{w \in W | \ \forall a' \in (\mathcal{A} - \{a\}) : \ U(a, w) > U(a', w)\} | \ a \in \mathcal{A}\}.$$

For each action $a \in \mathcal{A}$ we will denote the cell corresponding with a by a^*, and this, again, is exactly a hypothesis in the original set H. This shows that choosing a hypothesis can be thought of as a special kind of action.

But to show that a decision problem of the form $\langle P, H\rangle = \langle P, \mathcal{A}^*\rangle$ is a special case of a problem of the form $\langle P, U, \mathcal{A}\rangle$, we also have to eliminate the utility function in a natural way. The most natural way in which this can be done is to assume that for this case the utility function is the utility function of someone who cares about the truth, and nothing but the truth.

Suppose that there are only two units of utilities, u_1 and u_2, such that u_1 is strictly higher than u_2. In combination with the foregoing assumption this means that the actions taken in a world can be counted as being either wrong or right, i.e. having a utility of either 1 or 0; action a has utility 1 in a world iff hypothesis a^* is true in that world, and has utility 0 otherwise. Thus, the utility function is nothing else but a truth-value function. The utility value of choosing now is in these special circumstances the same as the probability value of the hypothesis with the highest utility:

$$\begin{aligned} UV_H(\text{Choose now}) &= max_{a\in\mathcal{A}} EU(a) \\ &= max_{a\in\mathcal{A}} \textstyle\sum_w P(w) \times U(a,w) \\ &= max_{a^*\in\mathcal{A}^*} [(\textstyle\sum_{w\in a^*} P(w) \times 1) + (\textstyle\sum_{w\notin a^*} P(w) \times 0)] \\ &= max_{a^*\in\mathcal{A}^*} \textstyle\sum_{w\in a^*} P(w) \\ &= max_{a^*\in\mathcal{A}^*} P(a^*). \end{aligned}$$

Now we can determine for each action $a \in \mathcal{A}$ its conditional expected utility with respect to new proposition C:

$$\begin{aligned} EU(a,C) &= \textstyle\sum_w P(w/C) \times U(a,w) \\ &= P(a^*/C). \end{aligned}$$

Thus, in these special cases the expected utility of action a after learning C is the same as the probability of a^* conditional on C. As a result it also follows that the action a which maximizes the expected utility conditional on learning new proposition C, is the proposition a^* which has the highest probability conditional on C. Now we can also determine the utility value of choosing after learning C:

$$\begin{aligned} UV_H(\text{Learn } C\text{, choose later}) &= max_{a\in\mathcal{A}} EU(a,C) \\ &= max_{a^*\in\mathcal{A}^*} P(a^*/C). \end{aligned}$$

In terms of this notion we can define a utility value of learning proposition C, $UV^*_{\mathcal{A}^*}(C)$, that slightly differs from the one defined in the previous section, $UV(C)$, in that according to the new function we immediately subtract the utility value of choosing now.

$$\begin{aligned} UV^*_{\mathcal{A}^*}(C) &= UV_H(\text{Learn } C\text{, choose later}) - UV_{\mathcal{A}^*}(\text{Choose now}) \\ &= max_{a\in\mathcal{A}} EU(a,C) - max_{a\in\mathcal{A}} EU(a) \\ &= max_{a^*\in\mathcal{A}^*} P(a^*/C) - max_{a^*\in\mathcal{A}^*} P(a^*). \end{aligned}$$

Thinking of $\mathcal{A}^*$ again as the set of hypothesis H, one can that $UV^*_H(C)$, in distinction with $UV_H(C)$, can have a negative value, but also a positive one in case C only strengthens the initially already preferred hypothesis.

Given our new definition of the utility value of *assertions*, $UV^*_H(C)$ it is, under the special circumstances sketched in this subsection, true that

$$UV^*_H(A) \geq UV^*_H(B) \quad \text{iff} \quad max_{h\in H} P(h/A) \geq max_{h\in H} P(h/B).$$

Thus, we have shown the utility value of an assertion is the larger, according to this measure function, the larger the probability of the hypothesis that has maximal posterior probability derived from it.

Notice that when $max_{h \in H} P(h/A) \geq max_{h \in H} P(h/B)$, it also holds that learning A reduces the entropy of H more than B does, in case H consists of 2 hypotheses, because in these cases $E_A(\{h, \neg h\}) \leq E_B(\{h, \neg h\})$. We can conclude that at least in these very special cases, utility values of assertions behave similar to their informational values: $UV_H^*(A) \geq UV_H^*(B)$ iff $IV_H(A) \geq IV_H(B)$. However, when H contains of more than 2 hypotheses the result doesn't go through anymore. The reason is, intuitively, that to determine $UV_H^*(A)$ we only look at the optimal hypothesis, while to determine $IV_H(A)$ we also look at the various sub-optimal hypotheses.

Let us now, finally, look at proposition C that completely resolves the issue. That is, let us look at the case where for each $h \in H$, it either holds that $C = h$, or $C \cap h = \emptyset$. Notice that in that case the value $max_{h \in H} P(h/C)$ will always be 1, and the utility value of C, $UV_H^*(C)$, depends only on the prior probability of h^*.[19] Let us now look at the question that completely corresponds with decision problem H, i.e. let us look at question H itself. We might evaluate the expected gain from this question, $EUV_H^*(H)$, by averaging over the corresponding expected values of the answers:

$$EUV_H^*(H) \quad = \quad \sum_{h \in H} P(h) \times UV_H^*(h),$$

because for each $h \in H$ it holds that $UV_H^*(h) = -P(h^*)$, we can conclude that for these special cases the expected gain from question H, $EUV_H^*(H)$, decreases as the prior probability of the least surprising message, i.e. h^*, increases.

5.3 Questioning procedures

We ended section 5.1 with a negative result: even if we can represent the actions of a decision problem by a set of propositions, i.e. by a partition like H, there still exists *in general* no connection between the *informational value* of a proposition A, $IV_H(A)$, and its *utility value*, $UV_H(A)$. Something similar is the case for questions: $EIV_Q(Q')$ is in general no special case of $EUV(Q)$. Let us assume that for every action a_i of our decision problem $\mathcal{A}$ there corresponds a set of worlds a_i^* in which a_i is the unique best action to perform. Assume $\mathcal{A} = \{a_1, ..., a_5\}$ and that the corresponding $\mathcal{A}^* = \{a_1^*, ..., a_5^*\}$ partitions the set of worlds compatible with what our agent believes. According to the prior probability function, all 'worlds' a_i^* are equally likely. Suppose, moreover that the utility function is as follows:

$$U(a_i, a_j^*) = 1, \text{ if } i = j, 0 \text{ otherwise.}$$

[19] Of course, this does not mean that the utility values of propositions are thus always independent of the propositions themselves. This is only the case when we only compare the utility values of different propositions that all *fully* resolve the issue.

In this case one should pick the a_i whose corresponding proposition has the maximal probability. Suppose we have two questions, $Q = \{q_1, q_2\}$ and $Q' = \{q'_1, q'_2\}$. The following table gives the probabilities of a^*_i given that we learn an answer to one of these questions:

$EIV_{\mathcal{A}^*}(Q) \neq EUV_{\mathcal{A}}(Q)$:

	q_1	q_2	q'_1	q'_2
a^*_1	0.4	0.3	0.5	0.2
a^*_2	0.3	0.4	0.2	0.5
a^*_3	0.1	0.1	0.1	0.1
a^*_4	0.1	0.1	0.1	0.1
a^*_5	0.1	0.1	0.1	0.1

Because $max_{a \in \mathcal{A}} EU(a, q'_i) = 0.5 > 0.4 = max_{a \in \mathcal{A}} EU(a, q_i)$, it is obviously the case that $EUV_{\mathcal{A}}(Q) < EUV_{\mathcal{A}}(Q')$. However, it turns out that $E_Q(\mathcal{A}^*) < E_{Q'}(\mathcal{A}^*)$ and thus that $EIV_{\mathcal{A}^*}(Q) > EIV_{\mathcal{A}^*}(Q')$. Thus, in general $EUV(Q)$ and $EIV(Q)$ do not behave monotone increasing with respect to one another.

However, as shown by Sneed (1967), in the following special case they do.[20] Suppose our agent wants to know which of the elements of $X_0 = \{x_1, ...x_N\}$ is true. Our agent may partition X_0 into $n \leq N$ disjoint, non-void subsets.

$$X^1_1, X^1_2, ..., X^1_n.$$

Now he is given the choice to pay a fee r to be told which member of the partition contains the true member of X. Say he is told X^1_1. If $N(X^1_1) \geq n$ he may then partition X^1_1 and pay r to be reliably told which member of this new partition contains the true member of X_0. The agent can go on in this way until every answer to a new question contains only an elements of one of the elements of X_0.

For any number n and N there is a finite number v of different questioning procedures of this sort that the agent could employ in attempting to discover which member of X is true. Call these n-ary questioning procedures for X at constant rate r. Let

$$QP_X = \{p_1, p_2, ..., p_v\}$$

be the mutually exclusive and jointly exhaustive propositions describing the employment of these different n-ary questioning procedures to discover which member of X_0 is true. The decision problem is now which questioning procedure to follow: $\mathcal{A} = QP_X$.

To determine the utility of new information with respect to a questioning procedure, we have to determine the utility of a questioning procedure for the remaining set of possibilities X'. We will assume that this depends completely on the *costs* of the questioning procedure, $C(p)$, and that this is measured in terms of the number of n-ary questions of procedure p that still has to be asked before the member of X_0 can be determined for certain. But this means that in the optimal

[20] For other special cases, see van Rooij (2004a).

case $max_{a \in \mathcal{A}} EU(a) = -min_{p \in QP_X} C(p) = -E^n(X)$ and $max_{a \in \mathcal{A}} EU(a, q) = -min_{p \in QP_X} C(p, q) = -E^n_q(X)$.[21]

To make life easier, we will make use of a *decision rule* that assigns a unique action to every possible answer to Q. Because Q is a partition, $Q(w)$ is simply the element of Q that has to be answered if w is the case. Now we can determine the utility value of the decision rule d with respect to question Q, $EU(d, Q) = \sum_w P(w) \times U(d(Q(w)), w)$. In terms of the utility of a decision rule, we can now show in a simple way that the expected utility value of a question with respect to the decision problem which questioning procedure to adopt if you want to know which member of X is true reduces to the expected informativity value of this question with respect to 'question' X:

$$\begin{aligned} EUV_{QP_X}(Q) &= max_d EU(d, Q) - max_{p \in QP_X} EU(p) \\ &= -min_d C(d, Q) - -min_{p \in QP_X} C(p) \\ &= -\sum_{q \in Q} P(q) \times E_q(X) - -E(X) \\ &= E(X) - E_Q(X) \\ &= EIV_X(Q). \end{aligned}$$

6 Conclusions and Outlook

In this paper I have shown how we can measure the usefulness, or *relevance*, of questions and answers using Stochastic Communication Theory, Inductive Logic and Statistical Decision Theory, and I have suggested that some of these measures are of greater value than others. In other papers I have used these notions for linguistic purposes to account for (i) the meaning of questions and assertions (van Rooij, 2003a,b); (ii) conversational implicatures (van Rooij, 2003c), and (iii) the licensing of polarity items (van Rooij, 2003d). In Van Rooij (2003a), for instance, I argue that measuring the relevance, or value, of questions and answers is of importance for linguistic theory, because it helps the answerer to determine what is actually expressed by an interrogative sentence, and the questioner to calculate which proposition is expressed by a declarative answer. What is expressed by interrogative and declaratively used sentences is very context-dependent, and depends heavily on the decision problem of the questioner. Assuming that both participants know what the decision problem of the questioner is, I propose that what is expressed by an interrogative sentence is that question that would be *most relevant* with respect to the questioner's decision problem.

In this paper I have implicitly assumed that the participants of a dialogue are always *cooperative*. In particular, that it can never do any harm for the questioner to make her decision problem public, and that the answerer will always help the questioner as much as he can to solve her decision problem by giving complete answers. Although cooperativity is standardly assumed in Gricean (1989) pragmatics, the participants of a dialogue do not always behave accordingly. It has been argued by Merin (1999), for instance, that for linguistic purposes we should

[21] From Shannon's noiseless coding theorem it follows that in general $E^n(X) \leq min_{p \in DP_X} C(p) < (E^n(X) + 1)$.

base our notion of *relevance* on the assumption that the two participants of a dialogue try to win an argument. Adopting Anscombre & Ducrot's (1983) conjecture that by making assertions we always want to *argue* for particular hypotheses, he suggests to measure the relevance of an assertion in terms of its argumentative function. Assuming that the two participants of a dialogue always argue for mutually exclusive hypotheses, he proposes to determine the relevance of assertion A with respect to hypothesis h in terms of Good's (1950) measure of the weight of evidence: $r_h(A) = \log(P(h/A)/P(\neg h/A))$. It is, perhaps, reassuring that adopting such a radically non-cooperative view on language use doesn't make our whole investigation useless. It turns out that $r_h(A)$ can also be defined as the difference between $\inf(A/\neg h)$ and $\inf(A/h)$, i.e. $r_h(A) = \inf(A/\neg h) - \inf(A/h)$, and it is easily seen that $r_h(A) = 0$ just in case the informative value of A with respect to *yes/no*-question $\{h, \neg h\}$, $IV_{\{h,\neg h\}}(A) = E(\{h, \neg h\}) - E_A(\{h, \neg h\})$, is 0, too. Thus, Merin (1999) takes a proposition to be a relevant argument with respect to an hypothesis, just in case we (in section 4) say it is relevant with respect to the corresponding *yes/no*-question. This doesn't mean that our notions of relevance are, thus, the same. It might well be that $r_h(A) < 0$ although $IV_{\{h,\neg h\}}(A) > 0$, and the other way around, due to the fact that Merin measures the relevance of assertions with respect single *hypotheses*, while we measure them with respect to *questions*, or *decision problems*.

Only very recently it has become clear that an analysis of relevance in terms of the hearer's decision problem is not quite appropriate to account for conversational implicatures: the speaker's beliefs and preferences should be taken into account as well. The proper way to do this would be to embed our information- and decision theoretic analyses into a more general game theoretic one. It would be beyond the scope of the present paper to discuss this embedding, though.

References

1. Anscombre, J.C. & O. Ducrot (1983), *L'Argumentation dans la Langue*, Brussels: Mardaga.
2. Bar-Hillel, Y. (1955), "An examination of Information Theory", in: *Philosophy of Science*, 22, pp. 86-105.
3. Bar-Hillel, Y. & R. Carnap (1953), "Semantic Information", *Proceedings of the Symposium on Applications of Communication Theory*, Butterworth Scientific Publications, London.
4. Blackwell, D. (1953), "Equivalent comparisons of experiments", *Ann. Math. Statist.*, 24: 265-272.
5. Carnap, R. (1950), *Logical Foundations of Probability*, University of Chicago Press. Chicago.
6. Dretske, F.I. (1981), *Knowledge and the Flow of Information*, MIT Press, Cambridge Massachusetts.
7. Good, I.J. (1950), *Probability and the Weighing of Evidence*, Griffin, London.
8. Good, I.J. (1966), "On the principle of total evidence", *British Journal for the Philosophy of Science*, 17: 319-321.
9. Grice, H.P. (1989), *Studies in the Way of Words*, Harvard University Press.

10. Groenendijk, J. (1999), "The logic of interrogation", in: T. Matthews & D.L. Strolowitch (eds.), *The Proceedings of the Ninth Conference on Semantics and Linguistic Theory*, CLC Publications, Stanford.
11. Groenendijk, J. and M. Stokhof (1984), *Studies in the Semantics of Questions and the Pragmatics of Answers*, Ph.D. thesis, University of Amsterdam.
12. Groenendijk, J. & M. Stokhof (1997), "Questions", in: J. van Benthem & A. ter Meulen (eds.), *Handbook of Logic and Language*, Elsevier, Amsterdam, pp. 1055-1124.
13. Hamblin, C. L. (1958), "Questions", *Australian Journal of Philosophy*, 36: 159-168.
14. Hartley, R.V.L. (1928), "Transmission of information", *Bell System Technical Journal*.
15. Jeffrey, R. (1965), *The Logic of Decision*, McGraw-Hill, New York.
16. Lehmann, E.L. (1959), *Testing Statistical Hypotheses*, New York: Wiley.
17. Lewis, D. (1988), "Relevant implication", *Theoria*, 54: 161-174.
18. Lindley, D. V. (1956), "On a measure of information provided by an experiment", *Ann. Math. Stat.*, 29: 986-1005.
19. Marschak, J. & M. Radner (1972), *Economic Theory of Teams*, Yale University Press, New Haven.
20. Marschak, J. (1974a), "Prior and posterior probabilities and semantic information", in: G. Menges (ed.), *Information, Inference and Decision*, D. Reidel, Dordrecht, pp. 167-180.
21. Marschak, J. (1974b), "Information, decision, and the scientist", in: C. Cherry (ed.), *Pragmatic Aspects of Human Communication*, D. Reidel, Dordrecht, pp. 145-178.
22. Merin, A. (1999), "Information, relevance, and social decisionmaking", in: L. Moss, J. Ginzburg, M. de Rijke (eds.), *Logic, Language, and Computation. Vol. 2*, Stanford, pp. 179-221.
23. Raiffa, H. & R. Schlaifer (1961), *Applied Statistical Decision Theory*, MIT Press, Cambridge, MA.
24. Rooij, R. van (2003a), "Questioning to resolve decision problems", *Linguistics and Philosophy*, 26: 727-763.
25. Rooij, R. van (2003b), "Asserting to resolve decision problems", *Journal of Pragmatics*, 35: 1161-1179.
26. Rooij, R. van (2003c), "Conversational implicatures and communication theory", in J. van Kuppevelt & R. Smith (eds.), *Current and New Directions in Discourse and Dialogue*, Dordrecht, Kluwer, pp. 283-303.
27. Rooij, R. van (200d), "Negative polarity items in questions", *Journal of Semantics*, 20: 239-273.
28. Rooij, R. van (2004a), "Utility, informativity and protocols', *Journal of Philosophical Logic*, 33: 389-419.
29. Rooij, R. van (2004b), "Utility of mention-some questions", *Research on Language and Computation*, 2: 401-416.
30. Rosenkrantz, R. (1970), "Experimentation as communication with nature", in: J. Hintikka & P. Suppes (eds.), *Information and Inference*, Reidel, Dordrecht, pp. 58-93.
31. Rosenkrantz, R. (1977), *Inference, Method and Decision*, D. Reidel, Dordrecht.
32. Savage, L.J. (1954), *The Foundations of Statistics*, New York: Wiley.
33. Shannon, C. (1948), "The Mathematical Theory of Communication", *Bell System Technical Journal*, 27: 379-423 and 623-656. The paper also appeared as the second chapter of C. Shannon & W. Weaver (eds.) (1949), *The Mathematical Theory of Communication*, University of Illinois Press.

34. Skyrms, B. (1990), *The Dynamics of Rational Deliberation*, Cambridge Mass.: Harvard University Press.
35. Sneed, J.S (1967), 'Entropy, information and decision', *Synthese*, 17: 392-407.
36. Stalnaker, R. (1978), "Assertion", in: P. Cole (ed.), *Syntax and Semantics, vol. 9: Pragmatics*, New York: Academic Press, pp. 315-332.
37. Wiener, N. (1950), *The human use of human beings*, Houghton Mifflin Company, Boston.