Utility, informativity and protocols

Robert van Rooy*
ILLC/University of Amsterdam
R.A.M.vanRooij@uva.nl

Abstract

Recently, natural language pragmatics started to make use of decision-, game-, and information theoretical tools to determine the usefulness of questions and assertions in a quantitative way. In the first part of this paper several of these notions are related with each other. It is shown that under particular natural assumptions the utility of questions and answers reduces to their informativity, and that the ordering relation induced by utility sometimes even reduces to the logical relation of entailment. The second part of the paper shows how different proposals (using either protocols or likelihood functions) to measure the relevance/utility of non-partitional questions come down to the same thing.

1 Introduction

Recently, quantitative notions of *relevance* have been proposed for questions and assertions and, based on an assumption of optimal relevance, used for linguistic applications. Bar-Hillel and Carnap's (1953) notion of informativity (their 'inf' value) has already been used in traditional pragmatics of natural language; Merin (1999a,b) proposes that the relevance of a proposition q should be seen as argumentative value with respect to an hypothesis h and measured by Good's (1950) notion of the weight of evidence, $log\frac{P(q/h)}{P(q/\neg h)}$; and van Rooy (2001, 2003a,b,c) measures the utility of interpretations of questions and answers in terms of the value of sample information and the reduction of entropy. These measures are used to determine what is actually expressed by an expression with an underspecified semantic meaning, and to account for certain pragmatic inferences, especially conversational implicatures (e.g. Parikh, 1992; Merin, 1999b; van Rooy, 2001, 2003a,c, and Schulz, 2002). In an inspiring paper, Bernardo (1979) showed that the expected informational value of an experiment/question is the natural special case of its expected utility value. The first contribution of this paper is to extend this investigation in several ways. The second contribution is to measure the relevance/utility of non-partitional questions, and show how different proposals (using either protocols or likelihood functions) come down to the same thing.

*This paper was originally presented at LOFT 5: Logic and the Foundations of the Theory of Games and Decisions, 2002, in Torino. I would like the participants of this conference for discussion and comments. Furthermore, I would like to thank Johan van Benthem, Reinhard Blutner, Balder ten Cate, Peter Grünwald, Katrin Schulz, and two anonymous reviewers for comments, and Darrin Hindsill for correcting my English. The research of this work is supported by a fellowship from the Royal Netherlands Academy of Arts and Sciences (KNAW), which is gratefully acknowledged.

2 Relating Utility with Informativity

In Merin (1999a,b) and van Rooy (2001, 2003a,b) it was argued that decision- and information theory can be used to define notions of relevance of questions and assertions, and that these notions can be used for linguistic applications. In this section I will first quickly rehearse some of these definitions and continue by showing that the information-theoretic measures can be seen as special cases of the decision-theoretic ones. In section 2.1, the utility of assertions and questions are defined making use of tools well-known from decision theory and information theory. The utility and informativity values are defined with respect to a background decision problem and background question, respectively. In section 2.2 it is shown that in natural special cases the utility of an assertion and a question reduces to its informativity (reduction of entropy), and that the ordering relation induced by utility sometimes even reduces to the logical relation of entailment. In section 2.2.1 I discuss Bernardo's (1979) proof that the utility of a question in natural cases comes down to its reduction of entropy. In 2.2.2 this argument is extended to that of assertions, and in 2.2.3 some qualitative relations between assertions and questions are shown to be special cases as well. Section 2.4 shows that the (expected) reduction of entropy of assertions and questions reduces in special cases to their *absolute* informativity. In section 2.4, finally, it is shown that some goal-directed notions of relevance used in pragmatic analyses of natural language can be seen as special cases of utility values of new information as well.

2.1 Utility and Informativity

The relevance of an assertion and a question is determined in terms of how far it helps to resolve a *decision problem.* Decision and Information theory represent these decision problems in different ways, however, giving rise to different kinds of measures.

2.1.1 Utility

In Bayesian decision theory, a *decision problem* of an agent can be modeled as a triple, $\langle P, U, A\rangle$, containing (i) the agent's probability function, P, (ii) her utility function, U, and (iii) the alternative actions she considers, A. If an agent facing such a decision problem has to choose now, she simply should choose the action with the highest expected utility, which can be thought of as the utility of choosing now:

$$\begin{aligned} UV(\text{Choose now}) &= max_i EU(a_i) \\ &= max_i \sum_w P(w) \times U(a_i, w) \end{aligned}$$

Suppose that before she chooses she learns proposition q. Then she will choose action a such that $EU(a,q) = max_i EU(a_i, q)$, where $EU(a,q) = \sum_w P(w/q) \times U(a_i, w)$. In terms of this notion we can determine the utility value of the assertion q. Referring to a^* as the action that has the highest expected utility according to the original decision problem, we can determine the *utility value* of new information q, $UV(q)$, as follows:

$$\begin{aligned} UV(q) &= UV(\text{Learn } q\text{, choose later}) - UV(\text{Learn } q\text{, choose } a^*) \\ &= max_i EU(a_i, q) - EU(a^*, q) \\ &= VSI(q) \end{aligned}$$

In statistical decision theory (cf. Raiffa & Schlaifer, 1961) this notion is known as the *value of sample information* q, $VSI(q)$. This value can never be negative. Now we can determine the *expected utility of a question* as the average expected utility of the answers, also known as the *expected value of sample information*, $EVSI(Q)$, which plays an important role in Bayesian statistics.

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

Notice that just like $VSI(q)$, also this value will never be negative. In fact, the value will be 0 only in case no answer to the question would have the result that the agent will change her mind which action to perform.

2.1.2 Informativity

If only truth is at stake, it seems natural to use Information theory (see Cover & Thomas (1991) for overview) to measure the quality of questions and answers. Suppose our agent wants to know the true answer to question Q'. The *entropy* of partition Q' with respect to probability function P, $E(Q')$, is defined as

$$E(Q') \quad = \quad \sum_{q' \in Q'} P(q') \times log_2 \frac{1}{P(q')}$$

The entropy of a question/partition measures the difficulty of the decision in determining which answer is true (measured in terms of the expected number of binary (yes/no) questions needed to determine the full answer). New information might *reduce* this *entropy*. Let's denote the entropy of Q' with respect to the probability function conditionalized by q by $E_q(Q')$. Now we will equate the *reduction* of entropy, $E(Q') - E_q(Q')$, with the *Informativity Value* of q with respect to decision problem Q', $IV_{Q'}(q)$:

$$IV_{Q'}(q) \quad = \quad E(Q') - E_q(Q')$$

In van Rooy (2001) it was suggested that this notion, used by Lindley (1956) already to measure the informativity value of a particular result of an experiment, is useful for linguistic applications. Because learning q might flatten the distribution of the probabilities of the elements of Q', $IV_{Q'}(q)$ might have a negative value. Still, if we define the informational value of question Q, the *Expected Informational Value* with respect to partition Q', $EIV_{Q'}(Q)$, as the *average* informativity value of the answers to Q, it turns out that the new notion will never have a negative value:

$$EIV_{Q'}(Q) \quad = \quad \textstyle\sum_{q \in Q} P(q) \times IV_{Q'}(q)$$

Our $EIV_{Q'}(Q)$ equals the well-known notion of *mutual information* between Q and Q', $I(Q, Q')$, and indeed it holds that $EIV_{Q'}(Q) = EIV_Q(Q')$.

2.2 Informativity as Utility

To determine the utility values we looked in section 2.1 both at the probabilities and at the utilities, while to determine the informativity values we looked only at the probabilities involved. This suggests that the informativity values of questions and assertions are *special cases* of the corresponding utility values. But how could this be? Whereas in the decision-theoretic analysis we look only at the *optimal* action to do, or hypothesis to bet on, in the information-theoretic analysis we take also all the sub-optimal hypotheses into account, and concentrate on the *whole* probability *distribution.* This suggests that to think of the information-theoretic values of questions and assertions as special cases of their decision-theoretic values, we should think of individual actions as actions that look at the whole probability distribution. We will do this first for questions and then for assertions.

2.2.1 Expected Informativity as Expected Utility

Before we will discuss the relation between the two measures with respect to questions, let us first have a closer look again at the *expected informational value* of question Q' with respect to partition, or decision problem Q, $EIV_Q(Q')$. We defined this measure in section 2.1.2 as the average entropy reduction of Q due to an answer to Q': $EIV_Q(Q') = \sum_{q \in Q'} P(q') \times IV_Q(q')$. But we noted that $EIV_Q(Q')$ is also known as *mutual information* between Q and Q', $I(Q, Q')$, a *symmetric* notion which is normally defined as follows:

$$I(Q, Q') \quad = \quad \sum_{q \in Q} \sum_{q' \in Q'} P(q \cap q') \times log \frac{P(q \cap q')}{P(q) \times P(q')} \quad = \quad EIV_Q(Q') \quad = \quad EIV_{Q'}(Q)$$

Decision theory deals with the logical process of rational decision making in situations of uncertainty. As such, the notion of 'rational belief' cannot be considered separately from the notion of 'rational action'. Beliefs are, actually or potentially, considered as inputs into the process of choosing some practical course of actions. But sometimes inferences, or statements of belief, may be regarded as *ends in themselves*, to be judged independently of any 'practical' decision problem.

Statements of belief can perhaps most obviously be considered as decision problems for weather forecasters, producers of information about the relevant meteorologic data. To find the connection between the utility and informational values of questions we will consider the problem how the forecaster's client might reward the forecaster in such a manner as to encourage him to make accurate estimates of probabilities.

Suppose that the weatherman's prior information and data yield the probability $pr = P(\text{Rain})$ that it will rain tomorrow. He then faces the decision problem of which probability pr' he should announce publicly in his evening forecast. This decision depends, of course, on his perceived utility function. *We* would like to be told the value pr actually indicated by all the data at hand, but if the weatherman acts in his own self-interest, he would, or so it seems, not be honest. From the belief that weather forecasters will incur more criticism from failing to predict a storm that arrives than from predicting one that fails to arrive, we suspect that they will systematically overstate the probability of bad weather, i.e., announce a value $pr' > pr$. Is it nevertheless possible to give the forecaster a utility environment that always will induce him to tell the truth?

As I understand from Good (1972), it appears that this problem was discussed already from 1950 on. Suppose that the possible weather conditions tomorrow can be described by partition Q, and that the forecaster's prior information and the data gives rise to a probability function P defined over Q. The forecaster's decision problem now is to choose which probability function over Q he should publicly announce. Thus, the alternative actions can be thought of as the announcements of various probability functions. In our decision-theoretic analysis above we have assumed that each action, a, has in each world, w, a certain utility, $U(a, w)$. The question that arises is whether we can we think of a system of rewards, i.e. a utility function, that guarantees honesty of the forecaster. Fortunately for us, this question has already been answered affirmatively by Bernardo (1979).[1] Recall from section 2.1 that the utility value of choosing now, UV(Choose now), was defined as the expected utility of the action with maximal expected utility. In case the decision problem is one of announcing a probability distribution over Q, this comes down to the following:

$$UV(\text{Choose now}) \quad = \quad max_i \sum_{q \in Q} P(q) \times U(pr_i, q)$$

What should the utility function be like? Well, we want our forecaster to be *honest*. This means that the utility value of choosing now should be maximal in case $pr = P$. Thus, $max_i \sum_{q \in Q} P(q) \times U(pr_i, q) = \sum_{q \in Q} P(q) \times U(P, q)$. This already constrains the way the utility function should be defined, or the system of rewards should be set up. Equally important, however, is that the reward should depend exclusively on his estimates of the event that *actually* will happen. For weather forecasts we are only interested in *the truth*: the value of a probability function, pr, is only to be assessed in terms of the probability it assigns to the actual outcome. So, if q turns out to be true, the agent should be rewarded only on the basis of his or her previous judgment about the plausibility of q. Bernardo (1979) shows that utility functions for the actions of announcing probability distributions defined over partition Q which satisfy the above two constraints must be of the form $U(pr, q) = (A \times log\ pr(q)) + B_q$, where $A > 0$ and the B_q's are arbitrary constants.[2]

Bernardo (1979) shows that the above logarithmic system of rewards does not only motivate the forecaster to be honest, but also stimulates him to acquire the maximal possible amount of information. It turns out that under this reward function, the expected *utility* value of a question will be equal to its expected *informational* value.

On the assumption that a^* is the action which has maximal utility with respect to the original decision problem, and that the optimal act to take (given q') for each q is to report the actual probability of q (conditional on q'), we can define the utility value of assertion/answer q' as follows (where $P_{q'}$ is P conditionalized by q'):

$$\begin{aligned} UV(q') &= \textstyle\sum_{q \in Q} P(q/q') \times [max_i EU(a_i, q) - EU(a^*, q)] \\ &= \textstyle\sum_{q \in Q} P(q/q') \times [EU(P_{q'}, q) - EU(P, q)] \end{aligned}$$

Using the fact that for the decision problems under consideration the utility functions must be of logarithmic form, the utility value of q' can also be described as follows:

$$UV(q') \quad = \quad \sum_{q \in Q} P(q/q') \times [((A \times logP(q/q')) + B_q) - ((A \times logP(q)) + B_q)]$$

[1] This result is closely related to Kelly's (1956) connection made between mutual information and the increase in the doubling rate (cf. Cover & Thomas, chapter 6). See also Bellman & Kalaba (1956) and especially the discussion in Good (1972).

[2] It is also demanded that Q contains more than two elements. Good (1972) claims that Gleason came already to the same conclusion in unpublished work.

Highschool mathematics teaches us that this is equal to

$$UV(q') \quad = \quad A \times \sum_{q \in Q} P(q/q') \times log \frac{P(q/q')}{P(q)}$$

Now we can show that the expected utility value of question Q', $EUV(Q')$, comes down (up to constant A) in these circumstances to its expected informational value with respect to Q, $EIV_Q(Q')$:

$$\begin{array}{lcl} EUV(Q') & = & \sum_{q' \in Q'} P(q') \times UV(q') \\ & = & \sum_{q' \in Q'} P(q') \times (\sum_{q \in Q} P(q/q') \times log \frac{P(q/q')}{P(q)}) \\ & = & \sum_{q' \in Q'} \sum_{q \in Q} P(q/q') \times P(q') \times log \frac{P(q/q') \times P(q')}{P(q) \times P(q')} \\ & = & \sum_{q' \in Q'} \sum_{q \in Q} P(q \cap q') \times log \frac{P(q \cap q')}{P(q) \times P(q')} \\ & = & EIV_Q(Q') \end{array}$$

We can conclude that the forecaster should do his best in his inquiries to acquire the maximum possible amount of information about Q, i.e., doing/asking the best possible experiment/question, so as to decrease that entropy.

We have seen that in inference reporting problems, the optimal solution is to choose the actual probability function, i.e. P. But what if the agent, for one reason or another, reports a function that is close to it, say pr? What is the utility difference, $UD_Q(P, pr)$, between them? Bernardo & Smith (1994) show that also this can be determined easily, and that it comes down to another interesting measure.

$$\begin{array}{lcl} UD_Q(P, pr) & = & \sum_{q \in Q} P(q) \times max_i U(pr_i, q) - U(pr, q) \\ & = & \sum_{q \in Q} P(q) \times [(A \times log P(q) + B_q) - (A \times log\ pr(q) + B_q)] \\ & = & A \times \sum_{q \in Q} P(q) \times log \frac{P(q)}{pr(q)} \\ & = & A \times D_Q(P||pr) \end{array}$$

The notion of $D_Q(P||pr)$ was introduced by Kullback & Leibler (1951) to measure the *divergence* between two probability functions. $D_Q(P||pr)$ is at least as great as 0, with equality iff $P = pr$. The measure is also known as the *relative entropy* between P and pr. Among the various measures between probability functions it is, arguably, the most interesting one, and so for at least two reasons. First, in distinction with other variation measures it picks out Jeffrey's (1965) extension of standard conditionalization as the unique rule that minimizes the variation distance between the prior and the posterior probability function that accommodates a non-perfect observation.[3] Second, in terms of Kullback & Leibler's divergence function we can describe the *mutual information* between two partitions in an appealing way (cf. Cover & Thomas). The mutual information is described as the divergence between the *actual* probability function P measuring the *joint entropy* of Q and Q', $E(Q \sqcap Q')$,[4] versus the probability function, pr, measuring this joint entropy in case Q and Q' were *independent* of, or *orthogonal* to, one another. This is due to the fact that in the latter case, for each element $q \cap q' \in Q \sqcap Q' :\ pr(q \cap q') = P(q) \times P(q')$.

[3]See Halpern (2003), proposition 3.11.2.

[4]Where $Q \sqcap Q' = \{q \cap q' : q \in Q\ \&\ q' \in Q'\ \&\ q \cap q' \neq \emptyset\}$.

$$\begin{aligned} D_{Q \sqcap Q'}(P||pr) &= \sum_{s \in Q \sqcap Q'} P(s) \times log \frac{P(s)}{pr(s)} \\ &= \sum_{q \in Q} \sum_{q' \in Q'} P(q \cap q') \times log \frac{P(q \cap q')}{P(q) \times P(q')} \\ &= I(Q, Q') \end{aligned}$$

For this reason, perhaps, the mutual entropy between Q and Q', $I(Q, Q')$, is also known as their *interaction* entropy. Notice that from the above equality we can conclude immediately to the well-known fact that the mutual information between Q and Q' is at least as great as 0, with equality iff Q and Q' are independent.

2.2.2 Informativity of a proposition as its Utility

In the previous subsection we followed Bernardo (1979) showing that in special circumstances the expected *utility* value of a *question* takes the form of its expected *informativity* value. This is nice and appealing, and you might wonder whether we could *extend* this result to *assertions*. Unfortunately, however, this is not possible for the following reason: as we saw in section 2.1, while the *utility value* of assertion q, $UV(q)$, can never be negative; for the *informativity value*, $IV_Q(q)$ this is possible. This gives rise to the following puzzle: Can we not define the utility value of a proposition in another way such that (i) the expected utility value of a question is still defined as the *average* utility value of its answers but with the result that it still equals the earlier defined $EUV(Q)$; and (ii) that in the special circumstances Bernardo concentrated on, the utility value comes down to its informativity value. In the following we show that this puzzle can be solved. We will denote the new utility value of assertion q by $UV^*(q)$ and define it as follows:

$$\begin{aligned} UV^*(q) &= UV(\text{Learn } q, \text{choose later}) - UV(\text{Choose now}) \\ &= max_i EU(a_i, q) - EU(a^*) \end{aligned}$$

Notice that in distinction with $UV(q)$, $UV^*(q)$ can be negative, just like the informativity value of the same proposition. But it has another appealing property too: even if learning q does not give rise to a change of opinion about which action to undertake, in distinction with $UV(q)$, $UV^*(q)$ will in these circumstances have a positive value in case q *strengthens* the choice that was already preferred. For this latter reason it seems that $UV^*(q)$ is a better measure for in how far q *resolves* the underlying decision problem than $UV(q)$ is.

Now we can define the expected utility value of question Q, $EUV^*(Q)$ as the average utility value of its answers:

$$EUV^*(Q) = \sum_{q \in Q} P(q) \times UV^*(q)$$

Fortunately for us we can show that this value is the appealing one we used before, $EUV^*(Q) = EUV(Q)$:

$$
\begin{array}{rcl}
EUV^*(Q) & = & \sum_{q\in Q} P(q) \times UV^*(q) \\
& = & \sum_{q\in Q} P(q) \times [UV(\text{Learn } q, \text{ choose later}) - UV(\text{Choose now})] \\
& = & \sum_{q\in Q} P(q) \times [max_i EU(a_i, q) - EU(a^*)] \\
& = & [\sum_{q\in Q} P(q) \times max_i EU(a_i, q)] - EU(a^*) \\
& = & [\sum_{q\in Q} P(q) \times max_i EU(a_i, q)] - [\sum_{q\in Q} P(q) \times EU(a^*, q)] \\
& = & \sum_{q\in Q} P(q) \times [max_i EU(a_i, q) - EU(a^*, q)] \\
& = & \sum_{q\in Q} P(q) \times UV(q) \\
& = & EUV(Q)
\end{array}
$$

By Bernardo (1979) we know that when utility functions for the actions of announcing probability distributions defined over partition Q which satisfies *honesty* and *actuality* must be of the form $U(pr, q) = (A \times log\ pr(q)) \times B_q$ where $A > 0$ and the B_q's are arbitrary constants. Ignoring these constants we can show that when we make use of our new utility values, not only for *questions* but also for *assertions* it holds that the utility value comes down to its informativity value: $UV(q') = IV(q')$. For the proof we assume that $EU^P_Q(pr, q') = \sum_{q\in Q} P(q/q') \times U(pr_{q'}, q)$, where by definition $\forall q : pr_{q'}(q) = pr(q/q')$.

$$
\begin{array}{rcl}
UV^*_{Q,P}(q') & = & max_i EU^P_Q(a_i, q') - EU^P_Q(a^*) \\
& = & max_i EU^P_Q(pr_i, q') - EU^P_Q(pr^*) \\
& = & [max_i \sum_{q\in Q} P(q/q') \times U^P(pr_{i,q'}, q)] - [\sum_{q\in Q} P(q) \times U^P(pr^*, q)] \\
& = & [\sum_{q\in Q} P(q/q') \times log_2 P_{q'}(q)] - [\sum_{q\in Q} P(q) \times log_2 P(q)] \\
& = & [\sum_{q\in Q} P(q/q') \times log_2 P(q/q')] - [\sum_{q\in Q} P(q) \times log_2 P(q)] \\
& = & [-(\sum_{q\in Q} P(q/q') \times -log_2 P(q/q'))] - [-(\sum_{q\in Q} P(q) \times -log_2 P(q))] \\
& = & -E^P_{q'}(Q) - -E^P(Q) \\
& = & E^P(Q) - E^P_{q'}(Q) \\
& = & IV^P_Q(q')
\end{array}
$$

Thus, when only truth is at stake, utility comes down to entropy reduction.

2.2.3 Qualitative variants

There are two reasons why q' could reduce Q's entropy more than q'' does, i.e., have a higher informativity value: (i) either because it *eliminates more cells* of the partition Q, or (ii) because it changes the probability distribution over the cells, i.e. it makes some cells of Q that have a positive probability more probable than others. Assume that we ignore the latter possibility, i.e., assume that when q' is learned, each element of Q consistent with q' has equal probability.[5] Then the above induced ordering relation comes down to the claim that q' is better to learn than proposition q'' just in case q' eliminates more cells of partition Q than q'' does. If this happens in *all* models, the ordering relation between propositions q' and q'' can be reduced even further:

$$IV_Q(q') > IV_Q(q'') \quad \text{iff} \quad \{q \in Q : \ q' \cap q \neq \emptyset\} \subset \{q \in Q : \ q'' \cap q \neq \emptyset\}$$

[5] Thus, for all $q \in Q$ it holds that $P(q/q') = \frac{1}{|\{q \in Q| \ q' \cap q \neq \emptyset\}|}$, where $|S|$ denotes the cardinality of set S.

It is worth remarking that in this way we have reduced the ordering of propositions in terms of entropy reduction to the *qualitative* ordering between answers that Groenendijk & Stokhof (1984) have proposed and applied to account for some linguistic phenomena. Note that when $Q = W$, the ordering relation $>$ comes down to (one-sided) entailment.

While Groenendijk & Stokhof proposed a qualitative analysis of a relation of 'comparative relevance' between answers, ten Cate (2000) has recently proposed a qualitative relation comparing the relevance of questions. He does this in terms of the composition relation, $\circ$, applied to questions. Assuming that Q and Q' denote the equivalence relations corresponding to their partitions, $Q \circ Q'$ denotes $\{\langle w, v\rangle : \exists w' : \langle w, w'\rangle \in Q \ \& \ \langle w', v\rangle \in Q'\}$. This relation corresponds with a function from elements of Q to propositions, or, alternatively, with the following cover: $\{\bigcup\{q' \in Q' : q \cap q' \neq \emptyset\} : q \in Q\}$. Now ten Cate defines that Q' is a more relevant question to ask than Q'', if we want to know which element of Q is true, $Q' >_Q Q''$ iff $Q' \circ Q \subset Q'' \circ Q$. Let us now denote the set of answers to Q that are compatible with q', i.e. $\{q \in Q : q \cap q' \neq \emptyset\}$, by q'_Q. Notice that $IV_Q(q') \geq IV_Q(q'')$ holds precisely iff $E_{q'}(Q) \geq E_{q''}(Q)$, which according to Groenendijk & Stokhof's qualitative ordering relation between answers in turn holds precisely iff $q'_Q \subseteq q''_Q$. We will show that something similar holds for ten Cate's qualitative comparative relation between questions. If we assume that all three questions, Q, Q' and Q'' correspond with partitions, $Q' \geq_Q Q''$ can be rewritten as follows:

$$\begin{array}{lll}
Q' \geq_Q Q'' & \text{iff} & Q' \circ Q \subseteq Q'' \circ Q \\
& \text{iff} & \forall p \in \{\bigcup\{q \in Q : q \cap q' \neq \emptyset\} : q' \in Q'\} : \\
& & \quad \exists p' \in \{\bigcup\{q \in Q : q \cap q'' \neq \emptyset\} : q'' \in Q''\} : p \subseteq p' \\
& \text{iff} & \forall p \in \{\bigcup q'_Q : q' \in Q'\} : \exists p' \in \{\bigcup q''_Q : q'' \in Q''\} : p \subseteq p' \\
& \text{iff} & \forall q' \in Q' : \exists q'' \in Q'' : \bigcup q'_Q \subseteq \bigcup q''_Q \\
& \text{iff} & \forall q' \in Q' : \exists q'' \in Q'' : q'_Q \subseteq q''_Q
\end{array}$$

Now the question arises with which of our quantitative notions does $Q' \circ Q$ correspond. The first guess would be $EIV_Q(Q')$, but that cannot be, because whereas $EIV_Q(Q')$ is symmetric, $Q' \circ Q$ is not. Instead, ten Cate's notion corresponds with that of the non-symmetric notion of *conditional entropy*: $E_{Q'}(Q)$, the entropy of Q after we learned the answer to Q': $\sum_{q' \in Q'} P(q') \times E_{q'}(Q)$. This notion depends only on the entropy of Q given an answer to the question Q', just like in the case of $Q' \circ Q$. In the qualitative variant, $E_{q'}(Q)$ depends only on q'_Q. In the comparative notion we exchange *average* values of the $E_{q'}(Q)$s and $E_{q''}(Q)$s, by a *quantificational* relation between the q'_Qs and q''_Qs.

2.3 Utility and Absolute informativity

2.3.1 Absolute informativity of propositions

In section 2 we have defined the informativity value of a proposition as the reduction of entropy of question, or decision problem, Q. The notion of entropy itself, however, seemed more like an *absolute* notion of informativity. Remember that to determine the entropy of a question we needed, for each of its answers q, to know the value of $log_2 \frac{1}{P(q)}$. This latter notion is sometimes called the *self-information* of q, and via Bar-Hillel & Carnap (1953) is also known as the (absolute) informativity value of q, denoted by $\inf(q)$:

$$\text{inf}(q) \quad = \quad log_2 \frac{1}{P(q)} \quad = \quad -log_2 P(q)$$

Thus, the entropy of a question Q is defined as the *average* informativity value of its answers. A notion of (absolute) informativity plays an important role in pragmatic analyses of appropriate natural language use. This notion is normally defined in terms of entailment, but according to Levinson (2000), among others, $\text{inf}(\cdot)$ plays a crucial role as well. It is obvious how the latter two notions hang together: proposition q entails proposition q' if the inf-value of q is *always*, i.e., regardless of the probabilities, at least as great as the inf-value of q', $\text{inf}(q) \geq \text{inf}(q')$. What is less obvious is that there also exists a relation between the inf-value of a proposition and the way it reduces entropy. In some specific circumstances, the informational value of a proposition with respect to an underlying question, $IV_Q(\cdot)$, and the absolute informativeness function, $\text{inf}(\cdot)$, behaves monotone increasing with respect to one another. This holds in particular in case the issue is 'what is the *world* like', and the worlds have *equal* probability. If W is the set of all worlds, it corresponds one-to-one with the finest-grained partition: the issue of how the world is like. To prove the above mentioned claim, we have to show that $IV_W(q) > IV_W(q')$ iff $\text{inf}(q) > \text{inf}(q')$. Because $\text{inf}(q) > \text{inf}(q')$ if and only if $P(q) < P(q')$ and because $IV_W(q)$ is defined as $E(W) - E_q(W)$, it is enough to show that $P(q) < P(q')$ if and only if $E_q(W) < E_{q'}(W)$. So, suppose that $P(q) < P(q')$ and that $\forall w \in W : \ P(w) = \frac{1}{|W|}$, where $|W|$ denotes the cardinality of W. Assuming that $P(q) = \sum_{w \in q} P(w)$, it follows that $P(q) = \frac{|q|}{|W|}$, and $P(q') = \frac{|q'|}{|W|}$. Now we want to show that $E_{q'}(W) > E_q(W)$.[6] This is done on the basis of the fact that in case $n > m$ it also holds that $log_2 n > log_2 m$.

$$\begin{array}{lcl}
E_{q'}(W) & = & \sum_w P(w/q') \times log_2 \frac{1}{P(w/q')} \\
& = & \sum_{q'} P(w/q') \times log_2 \frac{1}{P(w/q')} \\
& = & |q'| \times \frac{1}{|W|} \times \frac{|W|}{|q'|} \times log_2 [\frac{|q'|}{|W|} \times \frac{|W|}{1}] \\
& = & log_2 |q'| \\
& > & log_2 |q| = E_q(W).
\end{array}$$

In combination with the reduction of utility to our informativity value, $IV_Q(\cdot)$, we can conclude that when only truth is at stake, utility with respect to a very fine-grained, or delicate, issue, comes down to (absolute) informativity.

2.3.2 Entropy as Utility

Now the issue arises as to whether we can show something similar for questions. It turns out that we can now prove something more general. We will show that the expected informativity value of a question reduces to its entropy in case the underlying decision problem is what the world is like. In distinction, however, with the similar reduction in the previous section, in this case we don't need to make the additional assumption that the probabilities are equally distributed over the worlds. To prove the above claim, it turns out to be useful to make use of the notion of *conditional entropy.* We have defined $E_q(Q')$ as the conditional entropy of Q' after q is learned,

[6] Katrin Schulz sent me this proof to show that a claim of mine in another paper was true.

$$E_q(Q') = \sum_{q' \in Q'} P(q'/q) \times \inf(q'/q)$$

where $\inf(q'/q) = -log_2 P(q'/q)$. In terms of this notion we now define $E_Q(Q')$, the entropy of Q' conditional on Q. This is defined as the expected entropy of Q' conditional on learning an answer to question Q:

$$\begin{aligned} E_Q(Q') &= \textstyle\sum_{q \in Q} P(q) \times E_q(Q') \\ &= \textstyle\sum_{q \in Q} P(q) \times \sum_{q' \in Q'} P(q'/q) \times \inf\ (q'/q) \\ &= \textstyle\sum_{q \in Q} \sum_{q' \in Q'} P(q \wedge q') \times \inf\ (q'/q) \end{aligned}$$

Now we can define the expected informativity value of a question also in terms of this notion $E_Q(Q')$:

$$\begin{aligned} EIV_{Q'}(Q) &= \textstyle\sum_{q \in Q} P(q) \times IV_{Q'}(q) \\ &= \textstyle\sum_{q \in Q} P(q) \times [E(Q') - E_q(Q')] \\ &= E(Q') - [\textstyle\sum_{q \in Q} P(q) \times E_q(Q')] \\ &= E(Q') - E_Q(Q') \end{aligned}$$

The *joint entropy* of two questions Q and Q', $E(Q, Q')$, is defined as follows:

$$E(Q, Q') = \sum_{q \in Q} \sum_{q' \in Q'} P(q \cap q') \times \inf(q \cap q')$$

Now it can be shown (cf. Cover & Thomas) that for any two partitional questions Q and Q' of the same set of worlds, it holds that $E(Q, Q') - E(Q) = E_Q(Q')$:

$$\begin{aligned} E(Q, Q') - E(Q) &= -\textstyle\sum_{q \in Q} \sum_{q' \in Q'} P(q \wedge q') \times logP(q \wedge q') + \sum_{q \in Q} P(q) \times logP(q) \\ &= \textstyle\sum_q \sum_{q'} P(q \wedge q') \times logP(q) - \sum_q \sum_{q'} P(q \wedge q') \times logP(q \wedge q') \\ &= \textstyle\sum_{q \in Q} \sum_{q' \in Q'} P(q \wedge q') \times [logP(q) - logP(q \wedge q')] \\ &= \textstyle\sum_{q \in Q} \sum_{q' \in Q'} P(q \wedge q') \times log \frac{P(q)}{P(q \wedge q')} \\ &= \textstyle\sum_{q \in Q} \sum_{q' \in Q'} P(q \wedge q') \times \inf(q'/q) \\ &= E_Q(Q') \end{aligned}$$

Thus, $E_Q(Q') = E(Q, Q') - E(Q)$. It should be clear that the joint entropy of Q and Q' is equivalent to the entropy of $Q \sqcap Q'$, $E(Q \sqcap Q')$, where $Q \sqcap Q' \stackrel{def}{=} \{q \cap q' : q \in Q \ \& \ q' \in Q' \ \& \ q \cap q' \neq \emptyset\}$. In case $Q' \sqsubseteq Q$, i.e. if $\forall q' \in Q' : \exists q \in Q : q' \subseteq q$, it is clear that $(Q \sqcap Q') = Q'$, and thus also that $E(Q, Q') = E(Q')$. Now we can show that in case $Q' \sqsubseteq Q$, the expected information value of Q with respect to Q' equals the entropy of Q:

$$\begin{aligned} EIV_{Q'}(Q) &= E(Q') - E_Q(Q') \\ &= E(Q') + E(Q) - E(Q, Q') \\ &= E(Q') + E(Q) - E(Q'), \quad \text{if } Q' \sqsubseteq Q \\ &= E(Q), \quad \text{if } Q' \sqsubseteq Q \end{aligned}$$

Consider now W again, i.e. the set of all worlds, or its corresponding partition. Because W is the most fine-grained partition, it obviously holds that $W \sqsubseteq Q$ for any partitional question Q. It follows that the expected informativity value of Q with respect to W reduces to $E(Q)$: $EIV_W(Q) = E(Q)$. In combination with section 2.2 this means that if the decsion problem is to determine which world holds in as few yes/no questions as possible, the *entropy* of a question *measures* its *utility*. Because the entailment relation between questions is normally (e.g. Groenendijk & Stokhof, 1984) defined in terms of the fine-grainedness relation: $Q \models Q'$ iff $Q \sqsubseteq Q'$, and because $Q \sqsubseteq Q'$ iff for all P, $E^P(Q) \geq E^P(Q')$, we can conclude that the *entailment* relation between *questions* can be thought of as an *abstraction* from its *usefulness* relation.[7]

2.4 Argumentative Relevance

2.4.1 Utility and Argumentative Relevance

The information-theoretic way of determining the utility of an assertion of the previous subsection was based on the assumption that assertions are made in *cooperative* conversations to help in resolving another's decision problem. In such cooperative situations we might say that the speaker 'adopts' the decision problem of the other participant. But, of course, conversations need not be fully cooperative: sometimes a speaker just wants to change the common ground, represented by a shared probability function, in a particular direction. Suppose that our agent wants proposition h to be commonly accepted, or wants to be in a world where h holds. In that case we might determine the usefulness of proposition q as the difference between $P(h/q)$ and $p(h)$. Indeed, Merin (1999a,b) adopts something like $P(h/q) - P(h)$ to measure the *argumentative value* of q with respect to h, $AV_h(q)$, and uses it for linguistic applications.

Can we also think of this argumentative value of a proposition as a special case of its utility value?[8] Yes, we can, and the easiest way to see this is by making use of Jeffrey's (1965) decision theory, rather than of Savage's (1954), which we used until now. The difference between the two is that while Savage assumes that utility functions take action-world pairs as arguments, Jeffrey assumes these arguments simply to be worlds. As a result, Jeffrey can measure the value of an agent's current belief-desire state without considering real actions (in fact, he thinks of actions as propositions). In this framework, we can determine the expected utility of any proposition q as below.

$$EU(q) = \sum_{w \in q} P(w) \times U(w)$$

In particular, we can determine the utility of the *tautologous* proposition, $\top$, which measures the value of the current belief-desire state: $EU(\top) = \sum_{w \in \top} P(w) \times U(w) = \sum_w P(w) \times P(w)$. Let's denote this value by $EU(P, U)$. If $\langle P, U \rangle$ and $\langle P', U' \rangle$ are two different belief-desire states, $EU(P, U) > EU(P', U')$ means that the agent prefers to have the beliefs and desires as represented in $\langle P, U \rangle$ to the beliefs and desires in $\langle P', U' \rangle$. We will denote the probability function after conditionalizing P with q by P_q, and thus as-

[7] See section 3 for more discussion.

[8] Hintikka & Pietarinen (1966) have defended the measure $P(h/q) - P(h)$ in terms of decision theory as well. However, they did so under a different interpretation: For them the notion measures the expected utility of choosing h given evidence q. Thus, whereas we want to maximize $P(h/q_i) - P(h)$, they seek to maximize $P(h_i/q) - P(h_i)$. The defense I give is almost trivial and no claim of originality is made.

sume that $P_q(w) = P(w/q)$. Following section 2.2.2, we will measure the value of new information q, $UV^*(q)$, as follows:

$$UV^*(q) \quad = \quad EU(P_q, U) - EU(P, U)$$

The above representation of a belief-desire state and of the measure of new information are very general; in particular, what the utility function depends on is left open. If our agent has the desire to be in a world where h is true, or the goal to make that world actual where h is true, we can naturally assume that the utility function is defined as follows:

$$U(w) = 1 \quad \text{iff} \quad w \in h, 0 \text{ otherwise}$$

Notice that in this case the value $EU(P, U)$ reduces to the probability of h, $P(h)$,[9] and thus that the value of new information q reduces to the difference between $P(h/q)$ and $P(h)$, i.e. $UV^*(q) = P(h/q) - P(h)$.[10]

Most use of an argumentative notion of relevance to account for linguistic phenomena was made by Merin (1999a,b). To define the relevance of q given goal h, $r_h(q)$, he does not opt for measure $P(h/q) - P(h)$, however, but rather for $log\frac{P(q/h)}{P(q/\neg h)}$.[11]

$$r_h(q) \quad = \quad log\frac{P(q/h)}{P(q/\neg h)}$$

It is perhaps not very surprising anymore, but we can also think of this notion as a (natural) utility value of q. Assuming again that the agent always wants to choose for h and against $\neg h$, we might say that the the expected utility of the current belief-desire state, $EU(P, U)$, is $\frac{P(h)}{P(\neg h)}$.[12] Notice that this notion is higher or equal to 1 just in case $P(h) \geq P(\neg h)$, and that the situation gets better in case the number gets higher. Learning new information q might increase the difference between the probability of h and of $\neg h$, and thus also the fraction between them. The utility value of q might then be seen as

[9] $EU(P, U) = \sum_w P(w) \times U(w) = \sum_w P(w) \times (1, \text{ if } w \in h, 0 \text{ else}) = \sum_{w \in h} P(w) = P(h)$.

[10] In terms of Jeffrey's decision theory, we can also consider another special case: where $U(w) = logP(w)$. Now, the expected utility of belief-desire state $\langle P, U\rangle$ reduces to the negative entropy of partition W with respect to probability function P: $EU(P, U) = -E^P(W)$. Similarly, $EU(P_q, U) = -E^{P_q}(W)$, and thus $UV(q) = EU(P_q, U) - EU(P, U) = -E^{P_q}(W) - -E^P(W) = E^P(W) - E^{P_q}(W) = IV^P_W(q)$.

[11] Merin (1999b) notes that as far as ordinal preferences is concerned these notions, together with a number of others, are the same in the sense that $P(h/q) > / = / < 0$ iff $r_h(q) > / = / < 0$. Good's notion has some appealing properties, however, not shared by the standard notion of relevance. For instance, Merin (1997) shows that $r_h(\neg q) = -r_h(q)$, and that under natural conditions $r_h(q_1 \wedge q_2) = r_h(q_1) + r_h(q_2)$ and $r_h(q_1 \vee q_2) = \alpha r_h(q_1) + (1 - \alpha) r_h(q_2)$, for some $\alpha \in [0, 1]$. As it turns out, there are other relevance functions, e.g. Carnap's (1950) $r(h, q) = P(h \wedge q) - P(h) \times P(q)$ and $V(h, q) = log\frac{P(h \wedge q)}{P(h) \times P(q)}$, that have very similar properties, but I will not dwell upon this here.

[12] But how could the utility function look like on our assumption that $EU(P, U) = \sum_w P(w) \times U(w)$? The problem here is that the utility function works on worlds only, while it should, intuitively, take the whole probability function into account. There is, fortunately, a standard way to account for this: we assume that the probability function is *world*-dependent. Now we can define $U(w)$ as $log\frac{P_w(h)}{P_w(\neg h)}$. However, we wanted to derive that $log\frac{P(h)}{P(\neg h)}$ equals the *expected* utility, $\sum_w P(w) \times U(w)$, and not so much that it equals the utility/desirability in a particular world. But this can be accounted for by making the (standard) assumption that agents are *introspective*, meaning that in every world v in the support of P_w, $P_v = P_w$. Now it follows that for every such v: $U(v) = U(w)$, and thus also that $EU(P, U) = log\frac{P(h)}{P(\neg h)}$.

the difference between the expected utility value after and before this new information is received, and we end up with Merin's notion:[13]

$$
\begin{array}{rcl}
UV_h^*(q) & = & EU(P_q, U) - EU(P, U) \\
& = & log\frac{P(h/q)}{P(\neg h/q)} - log\frac{P(h)}{P(\neg h)} \\
& = & log\frac{P(h \wedge q)}{P(\neg h \wedge q)} - log\frac{P(h)}{P(\neg h)} \\
& = & logP(h \wedge q) - logP(\neg h \wedge q) - logP(h) - logP(\neg h) \\
& = & [logP(h \wedge q) - logP(h)] - [logP(\neg h \wedge q) - logP(\neg h)] \\
& = & log\frac{P(h \wedge q)}{P(h)} - log\frac{(\neg h \wedge q)}{P(\neg h)} \\
& = & logP(q/h) - logP(q/\neg h) \\
& = & log\frac{P(q/h)}{P(q/\neg h)} \\
& = & r_h(q)
\end{array}
$$

As noted by Merin (1999), this notion is also known as Good's (1950) *weight of evidence*, $W(h, q)$, who attributed it to Turing. According to Good (1972), however, the notion was used already by Peirce (1878), calling it the *weight of argument.* We have seen that in argumentative discourse studied by Merin, the notion is a rational measure.

2.4.2 Argumentative relevance and informativity

Somewhat counterintuitive, perhaps, but it can be shown easily that when A entails B, $A \models B$, it is still possible that A has a lower argumentative weight than B, i.e. there is an h such that $r_h(A) < r_h(B)$.[14] Because $r_h(A) \geq r_h(B)$ iff $AV_h(A) \geq AV_h(B)$ it is also possible that $A \models B$ but $AV_h(A) < AV_h(B)$. From this we can conclude that entailment and argumentative relevance don't behave monotone increasing with respect to one another. Something similar can be concluded for the relation between (absolute) informativity and argumentative weight: it doesn't hold in general that $\inf(A) \geq \inf(B)$ iff $r_h(A) \geq r_h(B)$.

However counterintuitive these facts might be, they should not really surprising you. We have shown above that both $AV_h(A)$ and $r_h(A)$ are special cases of $UV^*(A)$ and in section 2.2 that $UV^*(A)$ can have a negative value: although $A \models \top$, it can be that $UV^*(A) < UV^*(\top)$.

But perhaps there are natural circumstances under which argumentative relevance 'comes down' to informativity. As it turns out, this is indeed the case. In the previous section we said that proposition B has a positive argumentative value with respect to h, i.e. $AV_h(B) > 0$, just in case $P(h/B) > P(h)$. Notice that $P(h/B) > P(h)$ iff $P(h/B)/P(h) > 1$ iff $P(B/h)/P(B) > 1$. In fact, the measure $P(\cdot/h)/P(\cdot)$ behaves continuously monotone increasing with respect to our $AV_h(\cdot)$, meaning that if the one gets higher (lower), the other gets higher (lower) too. Notice that when $h \models B$, $P(B/h)/P(B) = \frac{1}{P(B)}$. The

[13]To continue our previous footnote, we have to assume now that learning q has the effect that for every world w, we go from probability function P_w to function $P_{q,w}$. In the spirit of Gerbrandy's (1999) update rule, we assume that v is in the support of $P_{q,w}$ if and only if (i) v is in the support of P_w and (ii) $v \in q$. This update rule has the desired effect that if a probability function is introspective, it remains introspective after conditionalization.

[14]See Merin (1999a). With Merin I think this is actually a strong point of relevance: it can account for the fact that you don't want to publically commit yourselve more than is needed.

function $\frac{1}{P(\cdot)}$, in turn, behaves continuously monotone increasing with respect to Bar-Hillel & Carnap's (1953) absolute informativity function, because $\inf(\cdot) = log\frac{1}{P(\cdot)}$. Thus, if h entails the arguments given, the measure $P(\cdot/h)/P(\cdot)$ behaves continuously monotone increasing with respect to $\inf(\cdot)$. But this means that in case h entails both A and B, it holds that $AV_h(A) \geq AV_h(B)$ iff $\inf(A) \geq \inf(B)$. The same observation can be made with respect to $r_h(\cdot)$ and $\inf(\cdot)$. Because $A \models B$ iff $\forall P : \inf^P(A) \geq \inf^P(B)$,[15] it also holds that in these circumstances $\forall P : r_h^P(A) \geq r_h^P(B)$ iff $A \models B$. Thus, whenever only truth is at stake, we can 'reduce' entailment to relevance in the following way:

$$A \models B \quad \text{iff} \quad \forall P, h: \text{ if } A, B \supseteq h, \text{then } r_h^P(A) \geq r_h^P(B)$$

We can conclude that in special circumstances maximizing argumentative relevance comes down to maximizing informativity.

3 Non-partitional questions

In terms of decision theory we can give a natural definition of the utility value of a question when the question is represented by a *partition*. To represent the meaning of a question by a partition is natural for questions like *Who came to the party?*, where the questioner wants to know the exact extension of the question-predicate 'Came to the party'. For other questions, however, a partition analysis is less natural. To give a resolving answer to the question *Who has got a light?*, one doesn't need to give a complete enumeration of all people that satisfy the question-predicate: naming one individual normally suffices. But this means that the meanings of several resolving answers might overlap: different resolving answers might be given in the same world, and the probability of getting an answer cannot be equated with the probability that the answer is true. If we assume that the meaning of a question is the set of its (possible) resolving answers, this means that this meaning in general need not partition the state space. In van Rooy (2004) it is argued that it is important to also determine the expected utility of non-partitional questions for linguistic applications. We will show that there are two ways to proceed, and will show that those two ways are actually the same. The first one uses *protocols*, the second *likelihood functions*.

3.1 Protocols

Until now we have assumed that incorporating new information goes by standard conditionalization. Problematic for this standard way of updating information states is, for instance, the well-known Monty Hall puzzle.[16] To account for the subtleties involved in such a puzzle, Halpern (2003) argues that conditionalization should be made dependent

[15] Where $\inf^P(A)$ denotes the informativity of A with respect to probability function P.

[16] As I was informed by Peter Grüwald, in statistics the question under which circumstances one can appropriately update a probability function by standard conditionalization after learning that a certain proposition is true is investigated extensively. As it turns out, in the context of *survival analysis*, Gill et al (1997) are able to provide a general characterization of these circumstances. In Grüwald & Halpern (2003) this characterization is generalized (to Jeffrey conditionalization, for instance) and given a procedural formulation.

on so-called *protocols*. As shown by van Rooy (2004), to determine the utility of non-partitional questions we can make use of such protocols as well. Let us assume that the answerer uses such a protocol or *answer rule*, a function f from information states to answers given. On the assumption that the answerer knows the true answer, we might say that it is a function from worlds to elements of Q. Let us temporarily assume that the questioner *does* know which f is in play. In that case, the utility of choosing after he learned answer q, $UV_f(\text{Learn } q, \text{choose later})$, should of course be determined as follows:

$$UV_f(\text{Learn } q, \text{choose later}) \quad = \quad max_i EU(a_i, f^{-1}(q))$$

In terms of this notion, we can now also define the utility of answer q, $UV_f(q)$:

$$\begin{array}{rcl} UV_f(q) & = & UV_f(\text{Learn } q, \text{choose later}) - EU(a^*, f^{-1}(q)) \\ & = & max_i EU(a_i, f^{-1}(q)) - EU(a^*, f^{-1}(q)) \end{array}$$

This value will never be negative. Notice that although question Q need not form a partition, it will be the case that with respect to each protocol f, the set $\{f^{-1}(q) | \ q \in Q\}$ will be a *partition*. Because of this, we can now determine the utility of question Q, $EUV_f(Q)$, as follows:

$$\begin{array}{rcl} EUV_f(Q) & = & \sum_{q \in Q} P_f(\text{get } q) \times UV_f(\text{get } q) \\ & = & \sum_{q \in Q} P(f^{-1}(q)) \times UV_f(q) \end{array}$$

Because no answer q can have a negative value, also the expected utility value of a question will never be negative.

The above determined utility of a question was based on the temporary assumption that the interrogator knows which protocol the answerer uses. But, of course, the interrogator himself doesn't know which one really will be used. The protocol actually used might be any rule f which satisfies *truthfulness*: $\forall w : w \in f(w)$. Let us denote the set of all protocols that satisfy truthfulness by F. The interrogator is uncertain which protocol will be used, but, let us assume, can quantify his uncertainty through a probability distribution. In terms of this we might determine the probability that the questioner gets answer q, $P_F(\text{get } q)$, as follows:

$$P_F(\text{get } q) \quad = \quad \sum_{f \in F} P(f) \times P(f^{-1}(q))$$

The utility value of getting answer q, $UV_F(\text{get } q)$, is determined similarly:

$$\begin{array}{rcl} UV_F(\text{get } q) & = & \sum_w \sum_{f \in F} P(f/q) \times P(w/f^{-1}(q)) \times [U(a_q, w) - U(a^*, w)] \\ & = & max_i \sum_{f \in F} P(f/q) \times [EU(a_i, f^{-1}(q)) - EU(a^*, f^{-1}(q))] \end{array}$$

Now we are ready to determine the expected utility of the question as expected:

$$EUV_F(Q) \quad = \quad \sum_{q \in Q} P_F(\text{get } q) \times UV_F(\text{get } q)$$

Notice that this definition is general and works also for partitional questions. In the latter case there will be only one possible protocol f that satisfies truthfulness. It follows that for any $q \in Q : f^{-1}(q) = q$, and thus that $\{f^{-1}(q) | q \in Q\} = Q$. Thus, in case Q itself is already represented by a partition, the above value of $EUV_F(Q)$ will be the same as the utility of the question as determined in section 2: $EUV(Q)$.

3.2 Likelihood

If answers overlap, there are worlds in which more than one answer might be given. The uncertainty which answer would be given in world w can be represented by a *likelihood function* η. For any world w and answer q to question Q, $\eta(q/w)$ denotes the likelihood that answer q will be given in world w. This likelihood function should obviously be constrained by the following conditions: for all q and w: $\eta(q/w) \geq 0$ and $\sum_{q \in Q} \eta(q/w) = 1$. Assuming that answers should be true, we also demand that if $\eta(q/w) > 0$, then $w \in q$. In terms of this likelihood function we can determine the probability of getting answer q and of the utility of getting answer q. First, the probability of getting answer q should be determined as

$$P_\eta(\text{get } q) \quad = \quad \sum_w P(w) \times \eta(q/w)$$

Before we can determine the utilities of answers and the question, we first have to know what the optimal action is after you learn the new information that q is true. After q is learned, the probability of w should not be determined with respect to the *prior* probability function P, but rather with respect to the *posterior* one. This posterior probability function, ρ, can be defined by means of Bayes' rule in terms of prior probability function P and likelihood function η as follows:

$$\rho(w/q) \quad = \quad \frac{P(w) \times \eta(q/w)}{\sum_v P(v) \times \eta(q/v)}$$

Notice that that the value of $\rho(w/q)$ need not be equal to $P((\{w\} \cap q)/q)$ and depends on the likelihood function η. Now we can determine what the optimal action is after you learned that q is true i.e. a_q:

$$a_q \text{ is the action which maximizes } \sum_w \rho(w/q) \times U(a_i, w)$$

If we denote by a^* the optimal action according to the original decision problem, we can determine the utility value of getting q, $UV(\text{get } q)$, in terms of $U(a_q, w)$ and the posterior probability function ρ as follows:

$$\begin{aligned} UV_\eta(\text{get } q) &= (\textstyle\sum_w \rho(w/q) \times U(a_q, w)) - (\sum_w \rho(w/q) \times U(a^*, w)) \\ &= \textstyle\sum_w \rho(w/q) \times (U(a_q, w) - U(a^*, w)) \end{aligned}$$

A non-trivial question corresponds to a non-trivial likelihood function. If η is a likelihood function it gives rise to the following set of propositions: $\{q \subseteq W | \exists w : \eta(q/w) > 0\}$. Notice that this comes down to a partition in the special case that for each w it holds that for each q, $\eta(q/w)$ is either 1 or 0.

Now we are ready to define the expected utility value of question Q, $EUV_\eta(Q)$, in the same way as Blackwell (1953), as the average utility value of the answers:

$$\begin{aligned} EUV_\eta(Q) &= \textstyle\sum_{q \in Q} P_\eta(\text{get } q) \times UV_\eta(\text{get } q) \\ &= \textstyle\sum_{q \in Q} \sum_w P(w) \times \eta(q/w) \times UV_\eta(\text{get } q) \end{aligned}$$

Notice that this definition is fully general. Indeed, it works both for partitional and non-partitional questions. In fact, when the likelihood function η is *trivial* in that for all w and q, $\eta(q/w)$ is either 1 or 0, everything is as before: $P(\text{get } q) = P(q \text{ is true})$ and $\rho(w/q)$ will just be the probability of w after conditionalizing the prior probability function P by q, i.e. $P(w/q)$.

3.3 Equivalence

Now we will show that the two ways of determining the expected value of questions come down to the same. The crucial assumption is that the likelihood function can be defined in terms of the protocols and the probabilities assigned to them:

$$\forall w, q: \ \eta(q/w) \quad = \quad \sum_{f \in F} P(f) \times \begin{cases} 1, & \text{if } w \in f^{-1}(q), \\ 0 & \text{otherwise} \end{cases}$$

We have seen above that the *posterior* probability function ρ depends on both the *prior* probability function and the likelihood function. If we translate the likelihood function as above, the posterior probability function ρ can be defined in terms of protocols as

$$\forall w, q: \ \rho(w/q) \quad = \quad \sum_{f \in F} P(f/q) \times P(w/f^{-1}(q))$$

In this formula, the conditional probability of f given q is determined via Bayes' rule as follows:

$$P(f/q) \quad = \quad \frac{P(f) \times P(q/f)}{\sum_{f \in F} P(f) \times P(q/f)} \quad = \quad \frac{P(f) \times P(f^{-1}(q))}{\sum_{f \in F} P(f) \times P(f^{-1}(q))}$$

Whether we determined the expected utility with respect to protocols, or with respect to a likelihood function, in both cases the definition went as follows:

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

To show that both analyses come down to the same, we have to show that $P_F(\text{get } q) = P_\eta(\text{get } q)$ and that $UV_F(\text{get } q) = UV_\eta(\text{get } q)$. The first can be shown as follows:

$$\begin{aligned} P_\eta(\text{get } q) \quad &= \quad \textstyle\sum_w P(w) \times \eta(q/w) \\ &= \quad \textstyle\sum_w P(w) \times \sum_{f \in F} P(f) \times \begin{cases} 1, & \text{if } w \in f^{-1}(q), \\ 0 & \text{otherwise} \end{cases} \\ &= \quad \textstyle\sum_{f \in F} P(f) \times P(f^{-1}(q)) \\ &= \quad P_F(\text{get } q) \end{aligned}$$

The second equivalence is equally straightforward:

$$\begin{aligned} UV_\eta(\text{get } q) \quad &= \quad \textstyle\sum_w \rho(w/q) \times [U(a_q, w) - U(a^*, w)] \\ &= \quad \textstyle\sum_w \sum_{f \in F} P(f/q) \times P(w/f^{-1}(q)) \times [U(a_q, w) - U(a^*, w)] \\ &= \quad UV_F(\text{get } q) \end{aligned}$$

3.4 Comparing utilities of questions and of information states

In van Rooy (2004) protocols were used to compare the utility of a *wh*-question under its partitional and its non-partitional reading, and it was argued that this comparison is important to determine which reading the *wh*-question actually will get. In many cases, the comparison between two (readings of) questions crucially involves the relevant probability and utility function of the underlying decision problem. Sometimes, however, we can say something more general. In this section I review two such general results and discuss their relation.

3.4.1 Blackwell's Theorem

The standard way to compare two partitions is in terms of the refinement relation $\sqsubseteq$. Partition Q is a refinement of Q', $Q \sqsubseteq Q'$ iff $\forall q \in Q : \exists q' \in Q' : q \subseteq q'$. And indeed, Groenendijk & Stokhof (1984) have defined the entailment relation between questions exactly in this way. In a very important paper, Blackwell (1953) has shown that the '$\sqsubseteq$' relation between two partitional questions corresponds with a natural relation between the expected utilities of these questions (where DP is a decision problem):

$$\text{For all partitional } Q \text{ and } Q' : Q \sqsubseteq Q' \quad \text{iff} \quad \forall DP: \ EUV_{DP}(Q) \geq EUV_{DP}(Q')$$

This very appealing result suggests that Groenendijk & Stokhof's *semantic* entailment relation between two questions is an abstraction from their *pragmatic* usefulness relation.

In fact, the above result is just a special case of a much more general theorem proved by Blackwell (1953). In the more general theorem he also compares two non-partitional questions, or better, their underlying likelihood functions, with one another.[17] For this Blackwell introduces a notion of *garbling*. We will say that two questions Q and Q', corresponding to likelihood functions η and η', respectively, can be related to each other by garbling G when for every $q \in Q$ and $q' \in Q'$, $G(q'/q)$ is a conditional probability meaning that $G(q'/q) \geq 0$ and $\sum_{q' \in Q'} G(q'/q) = 1$. Suppose we can find a garbling G that helps us to define the likelihood that answer q' will be given, $\sum_w P(w) \times \eta'(q'/w)$, in terms of likelihood function η as follows:

$$\eta'(q'/w) \quad = \quad \sum_{q \in Q} \eta(q/w) \times G(q'/q)$$

In that case also the likelihood function η' as a whole could be defined in terms of likelihood function η and garbling G. In matrix notation this is standardly stated by equation $\eta' = \eta G$.

Now we are ready to state **Blackwell's theorem** that compares two likelihood functions and thus their corresponding questions. If we denote the expected utility of likelihood function η with respect to decision problem DP by $EUV_{DP}(\eta)$, the theorem says the following:

$$(\forall DP: \ EUV_{DP}(\eta) \geq EUV_{DP}(\eta')) \quad \text{iff} \quad \eta' = \eta G, \text{ for some garbling } G$$

Notice that the fact mentioned earlier is indeed a special case of this theorem: the case where η and η' are noiseless and thus give rise to partitions. In those cases G is reduced to a many-to-one mapping from question Q corresponding with η to question Q' corresponding with η'. In that case Q' is *coarser* than Q because any element $q' \in Q'$ can be thought of as a collection of elements in Q: $\{q \in Q : \ G(q'/q) = 1\}$. Note that in these circumstances it holds that $G(q'/q) = 1$ iff $q \subseteq q'$, 0 otherwise.

[17] I became interested in comparing the two ways of determining utilities of non-partitional questions, because the utility of non-partitional questions in van Rooy (2004) was defined in terms of protocols while Blackwell's general result was defined in terms of likelihood functions.

3.4.2 Protocols to compare non-partitional information states

What we have called an answer rule, f, corresponds exactly with what in game theory (cf. Geanakoplos, 1994) is called a *possibility operator.* A possibility operator is supposed to represent the information processing capacity of an agent and is modeled as a function from worlds to propositions. Thus, it is modeled just like our protocols. If f is a possibility operator, for each w, $f(w)$ is interpreted as the collection of worlds the agent thinks are possible when the true state is w. In standard possible world terminology, $f(w)$ simply denotes the set of worlds *doxastically* accessible from w. Within game theory it is standardly assumed that a possibility operator f satisfies (i) *truthfulness*, $\forall w : w \in f(w)$; (ii) *positive introspection*, $\forall w, v \in f(w) : f(v) \subseteq f(w)$; and (iii) *negative introspection*: $\forall w, v \in f(w) : f(w) \subseteq f(v)$. These three conditions together assure that the possibility operator *partitions* the state space. In case f satisfies the following condition: $\forall v, w$; either $f(v) \cap f(w) = \emptyset$, or else $f(v) \subseteq f(w)$ or $f(w) \subseteq f(v)$, we say that it is *nested.* Notice that in case f gives rise to a partition it is nested, but that f can be nested without satisfying (modulo truthfulness and positive introspection) negative introspection.

Just as questions can be ordered by $\sqsubseteq$, possibility operators can be as well. We say that f is a *refinement* of g, $f \sqsubseteq g$, if for all w, $f(w) \subseteq g(w)$. Possibility operators are normally represented by functions that give rise to partitions. To account for things like speculative trade, however, game theorists have become interested recently in other functions too. Notice that for f to be a refinement of g, neither f nor g has to give rise to a partition. Observe also that because protocols/possibility operators give rise to questions, $Q_f = \{f(w)|w \in W\}$, we can also define the utility of functions f and g. Now we can state Geanakoplos' (1994) following theorem:

> If a possibility operator f satisfying *truthfulness*, *positive introspection* and *nestedness* is a refinement of a partition-inducing g, then $EUV(f) \geq EUV(g)$. Moreover, if f fails at least one of these conditions, then there exists a partition-inducing h that is less informative than f but has a higher utility.

You might wonder how this theorem could be true, for it seems to be *inconsistent* with Blackwell's theorem discussed above. It is easy to see that there exists a garbling for every likelihood function underlying question $Q = \{\{u, v\}, \{v, w\}\}$ to the trivial partition $Q' = \{\{u, v, w\}\}$, and the former is thus predicted to be at least as useful as the latter according to Blackwell's theorem. Still, no possibility operator underlying $\{\{u, v\}, \{v, w\}\}$ can satisfy nestedness,[18] which means that according to Geanakoplos' theorem $EUV(Q)$ can be lower than $EUV(Q')$. How could this be?

The reason is that Geanakoplos' analysis differs in two ways from ours. First, where we used a *set* of protocols to determine the utility of a question, Geanakoplos uses just one of them. Second, and crucially, he also determines the utility of choosing after learning a proposition q with respect to a protocol (or possibility operator) in a different way than we did. Recall that *we* defined the notion as follows:

$$UV_f(\text{Learn } q, \text{choose later}) \quad = \quad max_i EU(a_i, f^{-1}(q))$$

Geanakoplos (1994), however, makes use (implicitly) of the following notion:

[18] The two candidates are $\{\langle u, \{u, v\}\rangle, \langle v, \{v, w\}\rangle, \langle w, \{v, w\}\rangle\}$ and $\{\langle u, \{u, v\}\rangle, \langle v, \{u, v\}\rangle, \langle w, \{v, w\}\rangle\}$.

$$UV_f^*(\text{Learn } q, \text{ choose later}) \quad = \quad EU(a_q, f^{-1}(q))$$
where a_q is the action which maximizes $EU(a_i, q)$

Thus, UV_f^*(Learn q, choose later) denotes the expected utility of the action the 'naive' agent thinks is optimal after learning q itself, while our corresponding notion assumes that the agent knows the protocol 'behind' q and thus learns $f^{-1}(q)$.[19] This gives rise to the crucial difference that whereas our notion cannot be smaller than $EU(a^*, f^{-1}(q))$, Geanakoplos' notion can. It follows that for him also the expected utility of a non-partitional question can be negative, giving rise to the possibility that the protocol underlying a question like $\{\{u, v\}, \{v, w\}\}$ might have a lower utility value than the corresponding one underlying $\{\{u, v, w\}\}$.

3.5 Informativity of incomplete experiments

Suppose we want to measure the usefulness of a non-partitional question in case only truth is at stake. For that we use information theory again. Our first problem is to determine the entropy of a non-partitional question. The only obvious definition seems to be the following:

$$E(Q) \quad = \quad \sum_{q \in Q} P(\text{get } q) \times -logP(\text{get } q)$$

We have seen above that the notion $P(\text{get } q)$ can be defined equivalently in terms of protocols and likelihood functions. Assuming the latter, we see that the entropy of Q with respect to likelihood function η should be defined as follows:

$$E(\eta_Q) \quad = \quad \sum_{q \in Q}[\sum_w P(w) \times \eta(q/w) \times -log \sum_w (P(w) \times \eta(q/w))]$$

In terms of this notion we can now define the *informativity value* of proposition q' with respect to Q and η, $IV_{\eta_Q}(q') = E(\eta_Q) - E_{q'}(\eta_Q)$ and the *expected* informativity value of Q' with respect to Q and η, $EIV_{\eta_Q}(Q')$, and show that also these notions can be seen as special cases of the corresponding utility values in a similar way as we did in section 2.2. I will leave that to the interested reader. Let us see here, instead, how our 'obvious' notion of entropy of non-partitional questions relates with Rényi's (1961) analysis of *incomplete experiments*, the only other information-theoretic analysis I know of dealing with, what we call, non-partitional questions.[20] He extends Shannon's notion of entropy to incomplete experiments and gives an axiomatization of it. The Rényi-entropy of question/experiment Q, $RE(Q)$, is defined as follows:

$$RE(Q) \quad = \quad \frac{\sum_{q \in Q} P(q) \times -logP(q)}{\sum_{q \in Q} P(q)}$$

It is obvious that in case Q is a partition, $RE(Q)$ reduces to the standard notion of entropy. But how does $RE(Q)$ relate with our $E(\eta_Q)$? I don't see, to be honest, a

[19] In game theory this difference is known as *ex ante* versus *ex post* utility. In case f gives rise to a partition, the two always coincide.

[20] Thanks to Reinhard Blutner for pointing out the reference.

direct relation between the two measures. However, when we modify the Rényi-entropy as $RE'(Q)$, to be defined below, it is a natural special case of our own notion of entropy.

$$RE'(Q) \quad = \quad \sum_{q \in Q} \frac{P(q)}{\sum_{q \in Q} P(q)} \times -log \frac{P(q)}{\sum_{q \in Q} P(q)}$$

Our 'obvious' notion of entropy reduces to the modified Rényi-entropy if all answers to Q that are possible in a world are equally likely in this world, for in that case it holds that $P_\eta(\text{get } q) = \frac{P(q)}{\sum_{q \in Q} P(q)}$.

$$\begin{aligned} E(\eta_Q) &= \sum_{q \in Q} P_\eta(\text{get } q) \times -log_2 P_\eta(\text{get } q) \\ &= \sum_{q \in Q} [\frac{P(q)}{\sum_{q \in Q} P(q)} \times -log \frac{P(q)}{\sum_{q \in Q} P(q)}] \\ &= RE'(Q) \end{aligned}$$

4 Conclusion

In this paper I have shown how some notions used by linguists to measure the relevance of questions and answers can be reduced to more general Bayesian notions. Informativity defined as (expected) entropy reduction was seen to be the natural special case of (expected) utility value. Something similar holds for absolute informativity and some notions of argumentative value. Moreover, I have shown how the relevance of non-partitional questions can be determined, and how different ways of doing so come down to the same. Not being a mathematician, I have little doubt that the proofs made are not very deep. As a natural language theorist with a strong *functional* view on its use, however, I believe the results are of some interest.

References

[1] Bar-Hillel, Y. & R. Carnap (1953), 'Semantic information', *Proceedings of the Symposium on Applications of Communication Theory*, Butterworth Scientific Publications, London.

[2] Bellman, R. & R. Kalaba (1956), 'On the role of dynamic programming in Statistical Communication Theory', *IEEE; Transactions on Information Theory*, **3**: 197-203.

[3] Bernardo, J.M. (1979), 'Expected information as expected utility', *Ann. Statist.* **7**: 686-690.

[4] Bernardo, J.M. & A. Smith (1994), *Bayesian Theory*, John Wiley & Sons, LTD, Chichester.

[5] Blackwell, D. (1953), 'Equivalent comparisons of experiments', *Annals of Mathematical Statistics*, **24**: 265-72.

[6] Carnap, R. (1950), *Logical Foundations of Probability*, The University of Chicago Press, London.

[7] Cate, B. ten (2000), *Dynamic and epistemic semantics of questions: the logic of consultation*, Master thesis, Free University of Amsterdam.

[8] Cover, T. M. & J. A. Thomas (1991), *Elements of Information Theory*, Wiley series in telecommunications, Wiley, New York.

[9] Geanakoplos, J. (1994), 'Common knowledge', In R. Aumann & S. Hart (eds.), *Handbook of Game Theory, Volume 2*, pp. 1438-1496.

[10] Gerbrandy, J. (1999), *Bisimulations on Planet Kripke*, Ph.D. Dissertation, University of Amsterdam.

[11] Gil, R. D. & M. van der Laan & J. Robins (1997), 'Coarsening at random: Characterisations, conjectures and counter-examples', In: *Proceedings First Seattle Conference on Biostatistics*, pp. 255-294.

[12] Good, I.J. (1950), *Probability and the weighing of evidence*, London: Griffin.

[13] Good, I.J. (1972), 'Information, rewards, and quasi-utilities', in: Leach et al. (eds.), *Science, Decision, and Value*, D. Reidel Publishing Company, Dordrecht, pp. 115-127.

[14] Groenendijk, J. & M. Stokhof (1984), 'On the semantics of questions and the pragmatics of answers', In: F. Landman & F. Veltman (eds.), *Varieties of Formal Semantics*, Foris, Dordrecht.

[15] Grünwald, P.D. & J.Y. Halpern (2003), 'Updating Probabilities'. *Journal of Artificial Intelligence Research*, **19**: 243-278.

[16] Halpern, J. (2003), *Reasoning about Uncertainty*, MIT Press.

[17] Hintikka, J. & J. Pietarinen (1966), 'Semantic information and inductive logic', In J. Hintikka & P. Suppes (eds.) *Aspects of Inductive Logic*, North Holland.

[18] Jeffrey, R. (1965), *The Logic of Decision*, University of Chicago Press, Chicago.

[19] Kelly, J. (1956), 'A new interpretation of information rate', *Bell. Sys. Tech. Journal*, **35**: 917-926.

[20] Kullback, S. and R. Leibler (1951), 'On information and sufficiency', *Annals of Math. Statistics*, **22**: 79-86.

[21] Levinson, S.C. (2000), *Presumptive Meanings. The Theory of Generalized Conversational Implicatures*, Cambridge, Massachusetts: MIT Press.

[22] Merin, A. (1997), 'If all our arguments had to be conclusive, there would be few of them', *Arbeitspapiere des Sonderforschungsbereichs 340, No. 101*, Stuttgart.

[23] Merin, A. (1999a), *Die Relevanz der Relevanz: Fallstudie zur formalen Semantik der englischen Konjunktion* but, Habilitationschrift, University of Stuttgart.

[24] Merin, A. (1999b), 'Information, relevance, and social decisionmaking', In: L. Moss, J. Ginzburg, M. de Rijke (eds.), *Logic, Language, and Computation. Vol. 2*, Stanford.

[25] Lindley, D. V. (1956), 'On a measure of information provided by an experiment', *Ann. Math. Stat.*, **29**: 986-1005.

[26] Parikh, P. (1992), 'A game-theoretical account of implicature', in Y. Vardi (ed.), *Proceedings of the fourth TARK conference*, Monterey, California.

[27] Peirce, C.S. (1878), 'The probability of induction', *Popular Science Monthly*, reprinted in J. R. Newman (ed.), (1956), *The World of Mathematics*, Simon and Schuster, New york, pp. 1331-54.

[28] Raiffa, H. & R. Schlaifer (1961), *Applied Statistical Decision Theory*, MIT Press, Cambridge, MA.

[29] Rényi, A. (1961), 'On measures of entropy and information', in Neyman (ed.), *Proc. of the Fourth Symp. on Math. Stat. and Probability*, California.

[30] Rooy, van (2001), 'The relevance of communicative acts', In J. van Benthem (ed.), *Proceedings of the eighth TARK conference*, pp. 83-96.

[31] Rooy, R. van (2003a), 'Questioning to resolve decision problems', *Linguistics and Philosophy*, **27**: 493-527.

[32] Rooy, R. van (2003b), 'Negative polarity items in questions', *Journal of Semantics*, **20**: 239-273.

[33] Rooy, R. van (2003c), 'Quality and quantity of information exchange', *Journal of Logic, Language and Information*, **12**: 23-451.

[34] Rooy, R. van (2004), 'Utility of mention-some questions', *Research on Language and Computation*, **2**, vol. 3.

[35] Savage, L.J. (1954), *The Foundations of Statistics*, New York: Wiley.

[36] Schulz, K. (2002), *Relevanz und 'Quantity' Implikaturen*, Diplomarbeit, University of Stuttgart.

[37] Shannon, C. (1948), 'The mathematical theory of communication', *Bell System Technical Journal*, **27**: 379-423 and 623-656.