Pragmatic value and complex sentences

Robert van Rooij*

Abstract

In this paper we investigate to what extent it is possible to determine a reasonable default pragmatic value of complex sentences in a compositional manner, and – when combined with a Boolean semantics – to see under which conditions it gives rise to reasonable predictions. We will discuss several notions of pragmatic value, or relevance, and compare their behavior over complex sentences. We will see that although the goal-oriented notions of relevance give rise to the same ordering relations between propositions, the conditions under which they behave 'compositionally' vary significantly.

1 Introduction

The semantic meaning of a sentence is that aspect of the interpretation of the utterance of the sentence that can be determined compositionally via conventional rules from the conventional meanings of its parts. It is controversial to what the extent the interpretation of an utterance can be determined semantically. The traditional assumption is that for declarative sentences, semantics determines truth conditions. It is well-known, however, that in many contexts more is communicated with a declarative sentence than standard truth-conditional semantics predicts. Grice's (1967) pragmatic theory of conversational implicatures has been proposed to account for these extra non-truth-conditonal aspects of meaning. More recently (e.g. Sperber & Wilson 1986, Levinson 2000), pragmatics has been proposed to play a role in determining the truth conditions of a declarative sentence as well.

Within Grice's informal theory one assumption is of crucial importance: the assumption (combining his first subclause of the maxim of *Quantity* and his maxim of *Relevance*) that speakers provide as much information as they can as far as this is required, or relevant, in the conversation. But how do we determine whether a piece of information is relevant in a particular context, or more relevant than another piece of information? Can we define a natural notion of relevance that is applicable to pieces of information given in goal-oriented discourse? Over the years, numerous notions of 'relevance' have been proposed in the literature, varying from 'absolute' via 'comparative' to 'numerical' ones. In earlier work (e.g. Merin 1999a; van Rooij 2003, 2004) some of these notions have been compared with each other, and the main purpose of this paper is to extend these investigations. I will discuss some notions of relevance that I have not discussed before, and compare them mostly in terms of their behavior on complex sentences.

In standard model theoretic semantics it is assumed that the natural language connectives 'not', 'or', and 'and' should be represented by their logical analogues '$\neg$', '$\vee$', and '$\wedge$' and that these are interpreted in the standard Boolean way as complementation, union, and intersection, respectively. In terms of the Boolean operations we can then determine the semantic value (truth conditions) of a complex sentence (in as far as it depends on the connectives) in terms of the semantic values of its parts. In this paper I will adopt this assumption as well: the semantic value of a sentence is a proposition, and the semantic value of a complex sentence is determined in a Boolean manner.

*This paper was written as part of the 'Economics of Language' project, sponsored by the Dutch Organisation for Scientific Research (NWO), which is gratefully acknowledged.

Let us now assume that the pragmatic value of a sentence is its relevance. The interesting question that arises then is to what extent it is possible to do something similar here. Can we determine the pragmatic value of a complex sentence in terms of the pragmatic values of its parts? It should be clear that we should not be overly optimistic here. Consider, for instance, a sentence like *John is in the middle of the lake and he does not know how to swim.* This conjunctive sentence describes an emergency not present in either conjunct, and it is not easy to see how this pragmatic aspect of the conjunctive sentence can be determined from the pragmatic values of its parts. A second reason for being cautious is that the conjuncts of a conjunctive sentence can, for instance, stand in all kinds of logical relations to each other and this might give rise to strange predictions when we do not take into account the proposition expressed by the whole sentence. To meet these worries, our aim will be to investigate under which circumstances the proposed notions of 'pragmatic relevance' behave 'compositionally' in accordance with a Boolean analysis of semantic meaning. But what, then, could be the aim of determining the (default) pragmatic value in a compositional manner, if the result only corresponds to the actual pragmatic value of a complex sentence under these restricted circumstances? Why not first determine the semantic value of the complex sentence, and only then determine its relevance? With Merin (1993) we might propose that linguistic competence gives us rapidly the default pragmatic value of a complex sentence in terms of the values of its parts, but that this default value might be overridden when we have determined the truth conditions of the complex sentence. In any case, one aim of this paper is to investigate to what extent the pragmatic value can be determined compositionally in terms of a linear opereations, and – when combined with a Boolean semantics – to see under which conditions it gives rise to reasonable predictions.

The description I have given so far still leaves open various ways how to determine the relevance of a sentence, and, in turn, several ways how to determine the circumstances under which the pragmatic value of a complex sentence equals its default pragmatic value. Merin (1997, 1999a) favors Good's (1950) notion of *weight of evidence* as his notion of relevance and shows that under particular circumstances the (default) pragmatic value of a complex sentences can be determined as a linear function of the (default) pragmatic values of its parts. I believe that has begun an interesting project, and in this short paper I will slightly extend its scope. In particular, I will describe some other possible notions of pragmatic value, or relevance, and discuss to what extent, and under which circumstances, they behave in this way. I do so by stating a lot of (many would say) elementary facts whose proof in many cases is not my own. I can only hope that this collection of facts helps some readers to increase their understanding of the relation between several possible notions of pragmatic value of sentences in the same way as it helped me when writing this paper. Before I discuss some possible notions of pragmatic value, however, I will first give some motivation for why a default linear pragmatics makes sense in the first place.

2 Some motivations for a linear account

What would it mean to have a default compositional linear pragmatics for natural language? It would mean that the default pragmatic value of a complex sentence depends in a linear way on the pragmatic values of its parts. Let us assume that the pragmatic value of a proposition a is its usefulness, $U(a)$. Merin (1997) suggests that a straightforward linear pragmatics would then have it that the default pragmatic values of complex sentences are along the following lines: The usefulness of $\bar{a}$ would be the *opposite* of the usefulness of a with respect to some neutral point; the usefulness of $a \cap b$ would be, under some default conditions, the *sum* of the usefulness of a and the usefulness of b; and the usefulness of $a \cup b$ would – again under some default conditions – be a *convex combination* of the usefulness of a and of b. This gives rise to the following hypothesis.

Hypothesis 1 *Proposal for a default linear pragmatics*

$$\begin{array}{rcl} U(\bar{a}) & = & -U(a); \\ U(a \cap b); & = & U(a) + U(b); \\ U(a \cup b) & = & \alpha U(a) + (1-\alpha) U(b), \ \textit{with } 0 \leq \alpha \leq 1. \end{array}$$

We still need to know three things: (i) what does it mean to be the usefulness, or relevance, of a sentence?, (ii) what are the conditions under which the default value is the actual value, and, most importantly, perhaps, (iii) what is the motivation for treating *not* as *opposite*, *and* as *plus*, and *or* as *convex combination*? I will address the first two questions in the following sections. In this section I will only briefly address the third one, using arguments borrowed from Merin.[1]

The easiest 'utility' rule to motivate, perhaps, is the default additive behavior of **conjunctive** sentences. The first thing to notice is that in many linguistic contexts we can use 'and' and 'plus' interchangeably, and the latter obviously has addition as its meaning:

(1) a. Two and three is five.

b. Two plus three is five.

Second, suppose that (2) is given as the answer to the question who came to the party:

(2) At least John and Mary.

Intuitively, this answer is two times as useful as the single answers *(At least) John came* and *(At least) Mary came* separately. The final example not only suggests that *and* in some situations really has an additive meaning-component, but also gives an indication as to in which circumstances this default assumption cannot be made:

(3) a. John fell and (he) broke his arm.

b. John fell and he also broke his arm.

c. John fell AND he broke his arm.

Conjunctive sentence (3a) doesn't really have an additive reading. Such an assumption would ignore the most straightforward interpretation that John broke his arm *because* he fell. However, when we use *and also* or stressed *AND*, it is intuitively clear that the reader is informed that John's fall had nothing to do with why he broke his arm: the event denoted by the second conjunct occurred *independently* of the event denoted by the first. This suggests that conjunction behaves additively only in case the conjuncts are (taken to be) independent of one another.[2]

Of course, we have straightforward counterexamples to the assumption that conjunction behaves additively: 'A and A' isn't worth twice as much as 'A', and in the most straightforward reading of the following sentences, *and* has a non-distributive, and thus non-additive, reading:

(4) a. John and Mary (between them) have 5 euros.

b. Blücher and Wellington and Napolean fought against each other at Waterloo.

[1] For a much more serious defense of a linear approach towards natural language, the reader should consult the work of Merin himself.

[2] The reader might wonder, does the extra effort given in (3b) and (3c) compared to (3a) needed to 'derive' an additive reading not indicate that the default pragmatic value of a conjunctive sentence should *not* be that of addition? Perhaps, but, then, a sentence like (2) suggests otherwise.

The first problematic example just gives us another reason why independence matters for giving a default linear pragmatics, while Merin (1993, 1997) argues that examples (4a) and (4b) can be accounted for by a linear approach towards natural language meaning by providing an *indexical* meaning to *and* (and other connectives). In this paper, however, I will ignore this latter extension, and concentrate only on distributive uses of *and*.

There exists a substantive literature on the meaning and appropriate use of **disjunctive** sentences. What at least has to be explained is that it sometimes gives rise to an ignorance reading, as in (5a).

(5) a. John came to the party or Mary came to the party.

b. John or Mary came to the party and I don't know which.

In other contexts, *or* gives rise to a conjunctive and indifference reading: (6a) is most naturally interpreted as (6b):

(6) a. You may take an apple or a pear.

b. You may take an apple and you may take a pear, I don't care.

Merin (1997) proposes that the different readings come about due to the choice of α – the indexical element in the meaning of *or* – in hypothesis 1 that helps to determine the utility of a disjunctive sentence: in (5a) it is *nature* that is responsible for α, for (6a) the value of α is the *hearer's* responsibility. In the first case this results in the ignorance reading (5b), in the second it 'gives rise' to the free-choice inference (6b).[3]

Many people (see Horn, 1989) have argued that the use of a **negative** sentence is appropriate only in case its positive counterpart is expected, or highly relevant. This suggests that the use of negation has something to do with probability or expectation. The numerical pragmatic value associated in hypothesis 1 above with negation, however, is motivated by Merin (1997, 1999a,b) mainly from an *argumentative* perspective. He argues that the paradigmatic discourse situation is not one of cooperative information exchange, as is standardly assumed, but rather as one of explicit or tacit debate between two agents. One wants to argue for a certain 'goal' proposition g, and the other for its complement $\bar{g}$. Under this view of language use it is quite natural to assume that if proposition a favors g, its complement, $\bar{a}$, favors $\bar{g}$, and, consequently, has a negative utility with respect to g. Thus, whereas the conjunctive and disjunctive interpretations proposed in hypothesis 1 are mainly motivated by how we interpret sentences, for negation its pragmatic interpretation is rather based on a more general view of why we use language in the first place.

3 Assertability and informativity

3.1 Assertability

For '$-$' and '$+$' as used in hypothesis 1 to make sense, we have to assume that the pragmatic value of a sentence is a (real) number. The first candidate that comes to mind, of course, is to equate the pragmatic value of a sentence with its probability. And indeed, a number of authors (e.g. Adams, 1965) have proposed that this is a natural option because the subjective probability a speaker assigns to the semantic value of a sentence measures its *assertability* for that speaker,

[3]Of course, the free-choice inference only follows from the proposed linear analysis once a concrete, and non-standard, analysis of permission sentences is assumed. I have to admit, though, that I don't know exactly what such an analysis would look like.

or the *credibility* with which the speaker can assert the sentence.[4] Taking W to be a (finite) set of worlds, we call P a probability function in $W \rightarrow [0,1]$ iff $\sum_{w \in W} P(w) = 1$. In terms of this function, we can define for any proposition $a \subseteq W$ its probability as the sum of the probabilities of the worlds in a: $P(a) = \sum_{w \in a} P(w)$. In particular, we can do so for propositions denoted by 'complex' sentences.[5] Interesting for us is to see how we can determine the probability of a complex sentence in terms of the probabilities of its parts under particular circumstances. In order to do so, we first state the following (first) definition of independence.

Definition 1 *(Independence 1)* $[(a \perp b)_P]$ *iff* $P(a \cap b) = P(a) \times P(b)$.

Now, the following facts are in every text-book on probability theory.

Fact 1 *(probabilities of complex sentences)*

$$\begin{array}{rcl} P(\bar{a}) & = & 1 - P(a); \\ P(a \cup b) & = & P(a) + P(b), \ \textit{if } a \cap b = \emptyset; \\ P(a \cap b) & = & P(a) \times P(b), \ \textit{if } [(a \perp b)_P]. \end{array}$$

This result is interesting, but for union and intersection not exactly what we are looking for: it is neither clear how to explain the free-choice intuitions we have for *or*, nor why we associate *and* with additivity. However, the conditions under which *or* and *and* behave additive and multiplicative w.r.t. their probabilities remain important in this paper: the former is the case if the disjuncts are (mutually) *disjoint*; the latter if the conjuncts are *independent*. Remember that $P(a \cap b) = P(a) \times P(a/b) = P(a) \times P(b)$ iff $P(b/a) = P(b)$ (where $P(b/a)$ is defined as $\frac{P(a \cap b)}{P(a)}$), and thus that conjunction behaves multiplicatively if the conjuncts are conditionally independent of one another. As we will see below, by using a mathematical trick we can turn '$\times$' into '$+$' and thereby make conjunction additive.

3.2 Informativity and reduction of entropy

The well-known mathematical trick is to look at the (negative) logarithm of the probability of a proposition a, rather than at its probability. There is something more substantial behind this trick, however, because this new value measures naturally how surprising, or informative, learning a is. In fact, Bar-Hillel & Carnap (1953) defined the (absolute) 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:[6]

Definition 2 *(Absolute informativity)* $\inf(a) \quad = \quad \log_2 \ (1/P(a)) \quad = \quad -\log_2 \ P(a)$.

To explain the 'inf'-notion, let us consider a 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 $2^3 = 8$

[4]Except for conditionals. Adams (1965) has argued that the assertability of the conditional sentence should be equated with the conditional probability of the consequent given the antecedent. Lewis' (1973) well-known triviality result showed that there could be no (non-trivial) connective '$\Rightarrow$' with a context independent meaning such that the probability of a conditional sentence equals the conditional probability of the consequent given the antecedent. In this paper we will ignore the pragmatic value of conditional sentences, although there is a substantial literature dealing with exactly this issue.

[5]In the rest of this paper we assume that if the semantic values of sentences A and B are $[A] = a$ and $[B] = b$, respectively, the semantic values of $\neg A$, $A \vee B$, and $A \wedge B$ are $\bar{a}$, $a \cup b$, and $a \cap b$. We will sometimes confuse the distinction between sentences and their denotations and write, for instance, $\bar{a}$ for the sentence $\neg A$.

[6]Of course, we could measure the surprise value of a also as either $1 - P(a)$ or $1/P(a)$. Both measures give rise to the same ordering of propositions with respect to their informativity as $\inf(\cdot)$ does, but only the latter induces additivity of conjunction.

relevantly different states of the world, and assuming that it is considered to be equally likely for each 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 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 $M \cap S$, 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 $M \cap S$, $\inf(M \cap S)$, is 2. This, in fact holds in general: if a and b are independent, conjunction behaves informationally *additively*:

Fact 2 $\inf(a \cap b) = \inf(a) + \inf(b)$, *if* $[(a \perp b)_P]$.

For negative and disjunctive sentences, however, we cannot determine their informativity values straightforwardly in terms of the corresponding values of their parts. The informativity value of a proposition cannot be negative, and the informativity value of a disjunction in case the disjuncts are incompatible is not somewhere in between the informativity values of the disjuncts, as hypothesis 1 would have it, but rather (much) lower than the informativity value of each of its disjuncts.

It is easy to see 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. From this we can conclude that independent of the probability function, the trivial proposition $\top$ will always have an 'inf'-value of 0. 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.

The only other systematic fact that follows (as far as I can see) concerns the orderings between the informativity values of propositions:

Fact 3 $\inf(a) \gtrless \inf(b)$ *iff* $P(a) \lessgtr P(b)$, *and in particular* $\inf(\bar{a}) \gtrless \inf(a)$ *iff* $\inf(a) \lessgtr 1$ *iff* $P(a) \gtrless \frac{1}{2}$.

From fact 3 it immediately follows that 'inf' is monotonically increasing with respect to the entailment relation. That is, if $a \subseteq b$, it holds that $\inf(a) \geq \inf(b)$. Thus, the 'inf'-function is an extension of the partial ordering relation induced by the entailment relation. The entailment 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.

It is natural to assume that the informativity value of a proposition measures its pragmatic value in case the addressee wants to know exactly what the world looks like. In most, if not all cases, however, this is more than the addressee is really interested in. But, then, we can make the analysis more coarse-grained by defining the informativity value of a proposition in terms of the reduction of entropy of a partition. Suppose that our agent wants to know the true answer to question Q, represented by a partition of W. The *entropy* of partition Q with respect to probability function P, $E(Q)$, is defined as the average informativity of its answers.

Definition 3 $E(Q) \quad = \quad \sum_{q \in Q} P(q) \times \inf(q)$.

New information might *reduce* this *entropy.* Let's denote the entropy of Q with respect to the probability function P_a, which results from conditionalizing P by a, by $E_a(Q)$. Now we will equate the *reduction* of entropy, $E(Q) - E_a(Q)$, with the *informativity value* of a with respect to partition Q, $IV(Q, a)$:

Definition 4 $IV(Q, a) \quad = \quad E(Q) - E_a(Q)$.

Because learning a might flatten the distribution of the probabilities of the elements of Q, $IV(Q, a)$ might have a negative value. Of course, this won't happen if we assume that all elements of Q that have a positive probability according to both prior probability function P and posterior probability function P_a have an equal probability. In fact, in that case $IV(Q, a)$ behaves the same as $\inf(a)$, if we make the same assumptions with respect to the worlds consistent with P and P_a. And indeed, also $IV(Q, \cdot)$ has the desirable property that it behaves additively with respect to conjunction in case the conjuncts are probabilistically independent of each other. It is unclear, however, how the $IV(Q, \cdot)$-value of a negative or disjunctive sentence depends on the $IV(Q, \cdot)$-values of its parts.

Although we didn't gain much by looking at entropy reduction instead of absolute informativity to determine the pragmatic value of a complex sentence in terms of its parts, taking entropy reduction into account brings in two new ideas: (i) the pragmatic value of a sentence is a two-place relation and depends on a human concern, or a goal; and (ii) the pragmatic value of a sentence can be negative. As for (i), in this case, the goal is to know which proposition of the partition Q is true. But the goal can be much more general: it can be to solve the decision problem which action of a set $\{a_1, ..., a_n\}$ is the optimal action to perform, or, perhaps more abstractly, to be in a world in which a certain property holds. There exist various ways to define the utility value of a proposition in terms of the extent to which learning it helps to resolve a decision problem, and some of these notions have the $IV(Q, \cdot)$-value, and thus the $\inf(\cdot)$-value, as natural special cases.[7] For a systematic analysis of how the pragmatic value of a complex sentence depends on the pragmatic value of its parts, however, it turns out to be more useful to define the utility value of a proposition with respect to a concrete goal proposition. This is what Merin (1997, 1999a,b) proposed and also what I will do in the rest of the paper.

4 Standard statistical relevance

In statistics and decision theory we have a standard notion of relevance: with respect to goal proposition g, the usefulness of the acceptance of proposition a is given by the difference between the probability that g is true after and before proposition a is learned, i.e., $P(g/a)$ and $P(g)$.

Definition 5 *(Standard statistical relevance)* $R(g, a) = P(g/a) - P(g)$.

This notion is very intuitive, and in fact, it satisfies the intuitive requirements for being an adequate notion of relevance. For instance, proposition a is positively/negatively relevant with respect to g iff $R(g, a)$ is strictly higher/lower than 0. Obviously, when $a = \top$, no new information is received, and thus $R(g, \top) = 0$. In terms of proposition $\top$ we can now formulate a property of $R(\cdot, \cdot)$ that will actually be satisfied by all relevance functions to be discussed below. To state the following fact, I quantify implicitly over all propositions g and a, and all probability functions in terms of which R is defined.[8]

Fact 4 $R(g, a) \gtrless R(g, \top)$ *iff* $P(g \cap a) \gtrless P(g) \times P(a)$.

[7] For some discussion, see, among others, van Rooij (2003, 2004).

[8] I will do the same for all facts stated below, and I will normally not mention the implicit quantification anymore.

Proposition $\top$ is the prototypical *irrelevant* proposition, and thus we would like it to be the case that for all propositions g and a, if a is intuitively irrelevant to g, then $R(g,a) = R(g,\top) = 0$. Unfortunately, this doesn't follow immediately from the definition of relevance. Intuitively, $a = \bot$ is irrelevant as well. $R(g,\bot)$, however, is undefined, because $P(\bot) = 0$. More generally, we would like *irrelevance* to be a symmetric notion: $R(g,a) = 0$ if and only if $R(a,g) = 0$. But this doesn't hold if either g or a has a probability of 0. To account for irrelevance in a satisfying way, we therefore will follow Carnap (1950) in assuming that $R(g,a) = 0$ if $P(a) = 0$.[9] So, now we immediately account for the intuition that tautologies and contradictions are always irrelevant.

Fact 5 *If* $a = \top$ *or* $a = \bot$*, then* $R(g,a) = 0$.

The following fact will play an important role in proving certain other facts.

Fact 6 $R(g,a) \gtreqless 0$ *iff* $[P(g \cap a) \times P(\bar{g} \cap \bar{a})] - [(P(g \cap \bar{a}) \times P(\bar{g} \cap a)] \gtreqless 0$.

The proof of fact 6 is due to Carnap (1950) and makes use of the following abbreviations: $P(g \cap a) = k, P(\bar{g} \cap a) = l, P(g \cap \bar{a}) = m, P(\bar{g} \cap \bar{a}) = n$. Notice that $k + l + m + n = 1$. Now we prove fact 6 by proving that $k - (k + l) \times (k + m) = (k \times n) - (l \times m)$. This can be done as follows:

$$\begin{aligned} k - (k+l) \times (k+m) &= [k \times (k+l+m+n)] - [k \times (k+m) + l \times (k+m)] \\ &= k^2 + (k \times l) + (k \times m) + (k \times n)] - [k^2 + (k \times m) + (k \times l) + (l \times m)] \\ &= (k \times n) - (l \times m). \end{aligned}$$

To complete the proof, notice that $P(a \cap g) > P(g) \times P(a)$ iff $k - (k+l) \times (k+m) > 0$. Together with facts 4 and 5, it follows that fact 6 holds.

Fact 6 is very useful, because it immediately proves the following two facts already mentioned by Carnap (1950), which are interesting for linguistic applications if one looks at language use from an argumentative point of view:

Fact 7 $R(g,a) \gtreqless 0$ *iff* $R(\bar{g},a) \lesseqgtr 0$.

Fact 8 $R(g,a) \gtreqless 0$ *iff* $R(g,\bar{a}) \lesseqgtr 0$.

Thus, a is positively relevant with respect to goal proposition g iff a is negatively relevant with respect to goal proposition $\bar{g}$ iff $\bar{a}$ is negatively relevant with respect to goal g. The proof is nothing but applying fact 6 by exchanging g for $\bar{g}$ (for proving fact 7) and a for $\bar{a}$ (for proving fact 8). The following fact is much in the same spirit. It says that a is a better argument for g than b is iff b is a better argument for $\bar{g}$ than a is:

Fact 9 $R(g,a) \gtreqless R(g,b)$ *iff* $R(\bar{g},a) \lesseqgtr R(\bar{g},b)$.

$$\begin{aligned} R(g,a) \gtreqless R(g,b) \quad &\text{iff} \quad P(g/a) \gtreqless P(g/b) \\ &\text{iff} \quad P(\bar{g}/a) \lesseqgtr P(\bar{g}/b) \\ &\text{iff} \quad R(\bar{g},a) \lesseqgtr R(\bar{g},b). \end{aligned}$$

The above facts will hold for all relevance functions that will be discussed below. More surprisingly, perhaps, is that this also holds for the following fact. Merin (1997) shows for Good's (1950) notion of relevance that we can define classical negation in terms of the notions 'for' and 'against'. However, this already follows from the standard notion of relevance.[10]

[9] This assumption will be made for all other relevance functions discussed below as well.

[10] The reason behind this is that all relevance functions to be discussed give rise to the same ordering relation between propositions, they differ only in terms of the numbers assigned to the propositions that induce these orderings.

Fact 10 *$b = \bar{a}$ iff for all g and R, either one of a and b is an argument for g if the other is an argument against g.*

Fact 10 follows from fact 8 and the following fact:

Fact 11 *If for all $g : R(g,a) \gtrless 0$ iff $R(g,b) \lessgtr 0$, then $b = \bar{a}$.*

Our proof of fact 11 is exactly analogous to Merin's (1997) proof of a similar fact concerning Good's (1950) notion of relevance.[11] Suppose that $b \neq \bar{a}$, that is, it is not the case that ($a \cup b = W$ and $a \cap b = \emptyset$). There are 6 cases to consider: (1) $a \neq \emptyset \neq b$ and $a \cup b \neq W$; (2) $a \cap b \neq \emptyset, a \neq W \neq b$ & $a \cup b = W$; (3) $a = \emptyset \neq b$ and $a \cup b \neq W$; (4) $\emptyset \neq a \neq W = b$; (5) $a = b = \emptyset$; and (6) $a = b = W$. We prove for all cases that there are g and R such that the above claim is false.

(1)	$a \neq \emptyset \neq b$ and $a \cup b \neq W$. $g = a \cup b \Rightarrow \exists R : R(g,a) > 0$ & $R(g,b) > 0$.
(2)	$a \cap b \neq \emptyset, a \neq W \neq b$ & $a \cup b = W$. $g = a \cap b \;\Rightarrow\; \exists R : R(g,a) > 0$ & $R(g,b) = 0$.
(3)	$a = \emptyset \neq b$ and $a \cup b \neq W$. $g = b \;\Rightarrow\; \exists R : R(g,a) = 0$ & $R(g,b) > 0$.
(4)	$\emptyset \neq a \neq W = b$. $g = a \;\Rightarrow\; \exists R : R(g,a) > 0$ & $R(g,b) = 0$.
(5)	$a = b = \emptyset$. $R(g,a) = 0$ & $R(g,b) = 0$.
(6)	$a = b = W$. $R(g,a) = 0$ & $R(g,b) = 0$.

What about other complex sentences? In general, relevance gives rise in these cases to certain paradoxical predictions. For disjunctive sentences, for instance, it is possible that both a and b are positively relevant, but $a \cup b$ is not. It is easy to see, however, that under the assumption that $a \cap b = \emptyset$, this paradoxical prediction disappears.

Fact 12 *If $R(g,a) > 0$, $R(g,b) > 0$ and $a \cap b = \emptyset$, then $R(g, a \cup b) > 0$.*

$$\begin{array}{rcl} R(g, a \cup b) & = & P(g, a \cup b) - P(g) \\ & = & P(g/a) + P(g/b) - P(g), \text{ if } a \cap b = \emptyset \\ & \geq & [P(g/a) - P(g)] + [P(g/b) - P(g)] = R(g,a) + R(g,b). \end{array}$$

Also for conjunctive sentences we have a 'paradox'. It is possible that a and b are positively relevant to g, i.e., $R(g,a) > 0$ and $R(g,b) > 0$, but that $R(g, a \cap b) < 0$. The following fact states that under a certain (though perhaps not very natural) condition, this 'paradox' disappears.

Fact 13 *If $R(g,a) > 0, R(g,b) > 0$ and $P(g/a \cap b) \geq \min\{P(g/a), P(g/b)\}$, then $R(g, a \cap b) > 0$.*

$$\begin{array}{rcl} R(g, a \cap b) & = & P(g/a \cap b) - P(g) \\ & \geq & \min\{R(g,a), R(g,b)\} \\ & \geq 0 & \text{iff} \quad R(g,a) \geq 0 \;\&\; R(g,b) \geq 0. \end{array}$$

[11] Fact 19 will explain why we can take over his proof.

Though the disjointness assumption and the condition for conjunction saves us from paradoxes, this doesn't mean that the predicted relevance of disjunctive and conjunctive sentences is natural. As argued above, we would like the relevance of $a \cup b$ to be somewhere in between the relevance of a and b, i.e., we want it to be a convex combination of these relevances, and we desire the relevance of $a \cap b$ to be the relevance of a PLUS the relevance of b. As we will see in the rest of this paper, this is exactly what is predicted according to at least some notions of relevane.

5 Some other notions of relevance

Until now we have discussed the following relevance function:

Definition 5 (Standard statistical relevance) $R(g,a) = P(g/a) - P(g)$.

In this section I will define some other relevance functions that have been proposed in the literature, and show that they behave similar as far as the ordering they induce on propositions is concerned.[12]

Definition 6 *(Carnap's relevance)* $C(g,a) = P(g \cap a) - P(g) \times P(a)$.

Definition 7 *(Good's relevance)* $G(g,a) = \log \frac{P(a/g)}{P(a/\bar{g})} = \log \frac{P(g/a)}{P(\bar{g}/a)} - \log \frac{P(g)}{P(\bar{g})}$.

Definition 8 *(Keynes' relevance quotient)* $U(g,a) = \frac{P(g/a)}{P(g)}$.

Definition 9 *(Reduction of surprisal)* $V(g,a) = \log U(g,a)$.

We will now discuss some relations between those functions. The first observation shows the close connection between the standard $R(g,\cdot)$ and Carnap's $C(g,\cdot)$:

Fact 14 *$C(g,\cdot)$ is continuously monotone with respect to $R(g,\cdot)$, and thus in particular $C(g,a) \gtrless 0$ iff $R(g,a) \gtrless 0$.*

$$\begin{aligned} R(g,a) \gtrless R(g,b) \quad & \text{iff} \quad P(g/a) \gtrless P(g/b) \quad \text{iff} \quad \tfrac{P(a\cap g)}{P(a)} \gtrless \tfrac{P(b\cap g)}{P(b)} \quad \text{iff} \quad \tfrac{P(a\cap g)}{P(a)\times P(g)} \gtrless \tfrac{P(b\cap g)}{P(b)\times P(g)} \quad \text{iff} \\ & \text{iff} \quad [P(a \cap g) - (P(a) \times P(g))] \gtrless [P(b \cap g) - (P(b) \times P(g))] \quad \text{iff} \quad C(g,a) \gtrless C(g,b). \end{aligned}$$

We first extend this fact to $G(g,\cdot)$. Thus, we need to prove the following fact:

Fact 15 *$G(g,\cdot)$ is continuously monotone with respect to $R(g,\cdot)$ and in particular $G(g,a) \gtrless 0$ iff $R(g,a) \gtrless 0$.*

Because the logarithmic function is monotonically increasing with respect to its argument, we need to prove that $\frac{P(a/g)}{P(a/\bar{g})} \gtrless 0$ iff $P(g/a) \gtrless P(g)$ for the special case. For this, it is enough to prove the following (see also Merin, 1999b):

Fact 16 *$P(g/a) \gtrless P(g)$ iff $P(a/g) \gtrless P(a/\bar{g})$.*

[12] The various notions of relevance have been proposed mainly in the context of theories of confirmation within philosophy of science. Hintikka (1968), for instance, discusses our $R(g,a)$, $U(g,a)$, and $V(g,a)$ and calls them, trans-cont(a,g), trans-i(a,g), and trans-inf (a,g), respectively.

$$
\begin{array}{lll}
P(g/a) \gtrless P(g) & \text{iff} & \frac{P(g\cap a)}{P(a)} \gtrless P(g) \\
 & \text{iff} & \frac{P(g\cap a)}{P(g)} \gtrless P(a) \\
 & \text{iff} & P(a/g) \gtrless \alpha P(a/g) + (1-\alpha)P(a/\bar{g}), \text{ with } 0 \leq \alpha \leq 1 \\
 & \text{iff} & (1-\alpha)P(a/g) \gtrless (1-\alpha)P(a/\bar{g}) \\
 & \text{iff} & P(a/g) \gtrless P(a/\bar{g}).
\end{array}
$$

For the general case we show that $P(g/a) \gtrless P(g/b)$ iff $G(g,a) \gtrless G(g,b)$.

$$
\begin{array}{ll}
P(g/a) \gtrless P(g/b) & \text{iff} \quad \frac{P(g/a)}{1-P(g/a)} \gtrless \frac{P(g/b)}{1-P(g/b)} \quad \text{iff} \quad \log \frac{P(g/a)}{P(\bar{g}/a)} \gtrless \log \frac{P(g/b)}{P(\bar{g}/b)} \\
 & \text{iff} \quad G(g,a) \gtrless G(g,b) \quad (\text{because } G(g,a) = \log \frac{P(g/a)}{P(\bar{g}/a)} - \log \frac{P(g)}{P(\bar{g})}).
\end{array}
$$

To extend this to functions $U(g,\cdot)$ and $V(g,\cdot)$ it is enough to show the following:

Fact 17 $R(g,a) = P(g) \times [U(g,a) - 1]$.

$$
\begin{array}{lll}
R(g,a) & = & P(g/a) - P(g) \\
 & = & P(g) \times [\frac{P(g/a)}{P(g)} - 1] \\
 & = & P(g) \times [U(g,a) - 1].
\end{array}
$$

But this means that $U(g,\cdot)$ is continuously monotone with respect to $R(g,\cdot)$. Because $V(g,a) = \log U(g,a)$, this also means that $V(g,\cdot)$ is continuously monotone with respect to $R(g,\cdot)$. So we have established the following fact:

Fact 18 *$R(g,\cdot)$ is continuously monotone with respect to $U(g,\cdot)$ and $V(g,\cdot)$ and in particular $R(g,a) \gtrless 0$ iff $U(g,a) \gtrless 1$ iff $V(g,a) \gtrless 0$.*

The combination of facts 14, 15, and 18 establishes the fact that as far as ordinal relevance is concerned, it doesn't matter which of the relevance functions we use. They all give rise to the same relevance ordering between propositions.

Fact 19 *For all a, b, g and probability functions P i.t.o. which the relevance functions are defined:*

$$
\begin{array}{llll}
R(g,a) \gtrless 0 & \textit{iff} & C(g,a) \gtrless 0 & \textit{iff} \\
G(g,a) \gtrless 0 & \textit{iff} & U(g,a) \gtrless 1 & \textit{iff} \quad V(g,a) \gtrless 0.
\end{array}
$$

Fact 20 *For all a, b, g and probability functions P i.t.o. which the relevance functions are defined:*

$$
\begin{array}{llll}
R(g,a) \gtrless R(g,b) & \textit{iff} & C(g,a) \gtrless C(g,b) & \textit{iff} \\
G(g,a) \gtrless G(g,b) & \textit{iff} & U(g,a) \gtrless U(g,b) & \textit{iff} \quad V(g,a) \gtrless V(g,b).
\end{array}
$$

From these observations we can immediately derive the conclusion that facts 7, 8, 9, 10, and 11 established for $R(\cdot,\cdot)$ carry over to the other relevance functions:

Fact 21 *For all $f \in \{R, C, G, V\}$: $f(g,a) \gtrless 0$ iff $f(\bar{g},a) \lessgtr 0$ and $U(g,a) \gtrless 1$ iff $U(\bar{g},a) \lessgtr 1$.*

Fact 22 *For all $f \in \{R, C, G, V\}$: $f(g,a) \gtrless 0$ iff $f(g,\bar{a}) \lessgtr 0$ and $U(g,a) \gtrless 1$ iff $U(g,\bar{a}) \lessgtr 1$.*

Fact 23 *For all $f \in \{R, C, G, U, V\}$: $f(g,a) \gtrless f(g,b)$ iff $f(\bar{g},a) \lessgtr f(\bar{g},b)$.*

Fact 24 *$b = \bar{a}$ iff for all g and for all $f \in \{R, C, G, U, V\}$, any one of a and b is an argument for g if the other is an argument against g.*

Fact 25 *For all $f \in \{R, C, G, V\}$, If for all $g : f(g, a) \gtrless 0$ iff $f(g, b) \lessgtr 0$, then $b = \bar{a}$. Also, if for all $g : U(g, a) \gtrless 1$ iff $U(g, b) \lessgtr 1$, then $b = \bar{a}$.*

Having established the fact that as far as ordinal relevance is concerned, all the functions discussed here behave the same, we only have to discuss their numerical behavior. We will do so one by one.

6 Carnap's notion of relevance

We start with discussing Carnap's (1950) relevance function repeated below.

Definition 6 (Carnap's relevance) $C(g, a) = P(g \cap a) - P(g) \times P(a)$.

Notice that we can observe immediately from this definition that Carnap's notion of relevance is *symmetric*.

Fact 26 *$C(g, a) = C(a, g)$.*

In our proof of fact 6 above (due to Carnap himself) we already proved:

Fact 27 *$C(g, a) = [P(g \cap a) \times P(\bar{g} \cap \bar{a})] - [(P(g \cap \bar{a}) \times P(\bar{g} \cap a)]$.*

The following fact states that the positive relevance of $\bar{a}$ for g is the same as the negative relevance of a for g.[13]

Fact 28 *$C(g, \bar{a}) = -C(g, a)$.*

$$\begin{array}{rcl} C(g, \bar{a}) & = & [P(g \cap \bar{a}) \times P(\bar{g} \cap a)] - [P(g \cap a) \times P(\bar{g} \cap \bar{a})] \\ & = & [P(\bar{g} \cap a) \times P(g \cap \bar{a})] - [P(\bar{g} \cap \bar{a}) \times P(g \cap a)] \\ & = & -C(g, a). \end{array}$$

The following proposition (proved by Carnap) states a fact that Carnap's notion of relevance has in common with probability functions.

Fact 29 *$C(g, a \cup b) = C(g, a) + C(g, b) - C(g, a \cap b)$.*

$$\begin{array}{rcl} C(g, a \cup b) & = & P(g \cap (a \cup b)) - [P(g) \times P(a \cup b)] \\ & = & P((g \cap a) \cup (g \cap b)) - [P(g) \times [P(a) + P(b) - P(a \cap b)]] \\ & = & [P(g \cap a) + P(g \cap b) - P(g \cap a \cap b)] - [P(g) \times [P(a) + P(b) - P(a \cap b)]] \\ & = & [P(g \cap a) - P(g) \times P(a)] + [P(g \cap b) - (P(g) \times P(b))] - \\ & & \quad [P(g \cap a \cap b) - (P(g) \times P(a \cap b))] \\ & = & C(g, a) + C(g, b) - C(g, a \cap b). \end{array}$$

As an immediate corollary we receive additivity of the relevance of disjunction in case the disjuncts are taken to be mutually incompatible.

[13] All facts in this section were proved in Carnap (1950).

Fact 30 $C(g, a \cup b) = C(g, a) + C(g, b)$, *if* $P(a \cap b) = 0$.

Additivity of disjunction is not very remarkable, but it saves also Carnap's notion from the above mentioned paradox of disjunction. What is special about Carnap's notion of relevance, however, is that also conjunction is additive with respect to relevance. Notice that from fact 29 it immediately follows that $C(g, a \cap b) = C(g, a) + C(g, b) - C(g, a \cup b)$. But this means that if $P(a \cup b) = 1$, then $C(g, a \cup b) = 0$ and thus we have additivity for conjunction:

Fact 31 $C(g, a \cap b) = C(g, a) + C(g, b)$, *if* $P(a \cup b) = 1$ *and* $a \cap b \neq \emptyset$.

Of course, facts 30 and 31 can be generalized immediately to n-ary disjunctions and conjunctions respectively. Thus, if A is a collection of propositions that partitions $\bigcup A$, then $C(g, \bigcup A) = \sum_{a \in A} C(g, a)$. Similarly, if $P(\bigcup A) = 1$ and all elements in A are compatible, then $C(g, \bigcap A) = \sum_{a \in A} C(g, a)$.

To conclude, Carnap's notion of relevance comes closer to our ideal than the standard statistical notion, especially as far as conjunctive sentences is concerned. However, it cannot explain the free-choice effect of disjunctive sentences.

7 Good's notion of relevance

Merin (1997) makes use of Good's (1950) notion of relevance defined as follows:[14]

Definition 7 (Good's relevance) $G(g, a) = \log \frac{P(a/g)}{P(a/\bar{g})}$.

The following fact follows immediately (because $\log \frac{P(a/g)}{P(a/\bar{g})} = \log P(a/g) - \log P(a/\bar{g})$):

Fact 32 $G(\bar{g}, a) = -G(g, a)$.

Good's notion of relevance is additive for conjunction under a certain condition. The condition is the following notion of independence:

Definition 10 *(Independence 2) a and b are independent conditionally on g and $\bar{g}$ with respect to P,* $[(a \perp b | \pm g)_P]$ *iff* $P(a \cap b/g) = P(a/g) \times P(b/g)$ *and* $P(a \cap b/\bar{g}) = P(a/\bar{g}) \times P(b/\bar{g})$.

With the help of this definition, Merin (1997) shows that the following fact holds:

Fact 33 *(Additivity of conjunction): If* $[(a \perp b | \pm g)_P]$, *then* $G(g, a \cap b) = G(g, a) + G(g, b)$.

So far, there is not much that makes Good's notion preferable to Carnap's notion, although the notion of independence required for additivity differs. However, Good's notion of relevance behaves more nicely with respect to *disjunctive* sentences than Carnap's notion does. Merin (1997) shows that, under natural conditions, the weight of argument of a disjunction is a convex combination of the weight of arguments of its disjuncts:

Fact 34 *If* $P(a \cap b) = 0$, *then there is some* $\alpha \in [0, 1]$ *s.t.* $G(g, a \cup b) = \alpha G(g, a) + (1 - \alpha) G(g, b)$.

The above facts show that Merin's proposal to adopt Good's notion of weight of argument as his notion to measure relevance is natural because this notion behaves in many respects just as he wants it to do. However, this does not rule out that there exist other notions of relevance with perhaps equally appealing properties. In section 9 I will define a notion of relevance that will display almost exactly the same behavior for complex sentences as the notion used by Merin. To prepare the way for this new notion, however, it will prove to be useful first to discuss an old notion of relevance due to the economist Keynes.

[14] Good attributes the notion to Turing, although apparently the notion already goes back to Peirce (1878).

8 Keynes' relevance quotient

Keynes (1921) proposed the following notion of relevance as an improvement of the standard $R(\cdot,\cdot)$ discussed in section 4:

Definition 8 (Keynes' relevance quotient) $U(g,a) = \frac{P(g/a)}{P(g)}$.

In contrast to the standard notion and to Good's notion of relevance, but in agreement with Carnap's notion, Keynes' notion of relevance is *symmetric.*

Fact 35 *$U(g,a) = U(a,g)$.*

The proof is straightforward and follows almost immediately from the definition of $U(g,a)$:

$$\begin{array}{rclcl} U(g,a) & = & \frac{P(g/a)}{P(g)} & = & \frac{1}{P(g)} \times \frac{P(g\cap a)}{P(a)} \\ & = & \frac{P(g\cap a)}{P(g)\times P(a)} & = & \frac{1}{P(a)} \times \frac{P(a\cap g)}{P(g)} \\ & = & \frac{P(a/g)}{P(a)} & = & U(a,g). \end{array}$$

Relevance functions typically give rise to *paradoxes* if complex sentences are considered. We have seen already in previous sections that if natural conditions apply, these paradoxes disappear. This is also true for Keynes' notion. In fact, for Keynes' notion something stronger holds: $U(g,a)$ behaves as a *cardinal utility* as defined in Jeffrey (1965).

Fact 36 *If $P(a\cap b) = 0$, then $U(g,a\cup b)$ is a convex combination of $U(g,a)$ and $U(g,b)$.*

The following proof is due to Merin (1997).

$$\begin{array}{rcl} U(g,a\cup b) & = & \frac{P(g\cap(a\cup b))}{P(g)\times P(a\cup b)} \\ & = & \frac{P(g\cap a)+P(g\cap b)}{[P(g)\times P(a)]+[P(g)\times P(b)]}, \text{because } P(a\cap b)=0 \\ & = & \frac{P(g\cap a)+P(g\cap b)}{[P(a)+P(b)]\times P(g)} \\ & = & \frac{1}{P(a)+P(b)} \times [(P(a)\times \frac{P(g\cap a)}{P(a)\times P(g)}) + (P(b)\times \frac{P(g\cap b)}{P(b)\times P(g)})] \\ & = & \frac{[P(a)\times U(g,a)]+[P(b)\times U(g,b)]}{P(a)+P(b)}, \text{ is Jeffrey (1965) desirability} \\ & = & [\frac{P(a)}{P(a\cup b)}\times U(g,a)] + [\frac{P(b)}{P(a\cup b)}\times U(g,b)], \text{ is mixture} \\ & = & \alpha U(g,a) + (1-\alpha)U(g,b), \text{ where } \alpha = \frac{P(a)}{P(a\cup b)}. \end{array}$$

With respect to conjunction we can prove that it behaves multiplicational under some natural independence conditions. The natural condition is independence conditionally on g:

Definition 11 *(Independence 3) a and b are independent conditionally on g with respect to P, $[(a\perp b|g)_P]$ iff $P(a\cap b/g) = P(a/g)\times P(b/g)$.*

Now we can state the behavior of conjunction under this and the standard independence condition:

Fact 37 *If $[(a\perp b)_P]$ and $[(a\perp b)_P|g]$, then $U(g,a\cap b) = U(g,a)\times U(g,b)$.*

$$\begin{array}{rcl} U(g,a\cap b) & = & \frac{P(g/a\cap b)}{P(g)} \\ & = & \frac{P(a\cap b/g)}{P(a\cap b)} \\ & = & \frac{P(a\cap b/g)}{P(a)\times P(b)}, \text{if } P(a\cap b) = P(a)\times P(b) \\ & = & \frac{P(a/g)\times P(b/g)}{P(a)\times P(b)}, \text{if } P(a\cap b/g) = P(a/g)\times P(b/g) \\ & = & \frac{P(a/g)}{P(a)}\times\frac{P(b/g)}{P(b)}, \text{ under independence conditions} \\ & = & U(g,a)\times U(g,b), \text{ under independence conditions.} \end{array}$$

Notice that this also rules out the paradox for conjunction: if a and b are positively relevant, $a \cap b$ must be so as well.

9 Relevance as reduction of surprisal

We saw above that if relevance is measured by $U(g, \cdot)$, the relevance of a disjunctive sentence is appealingly enough a convex combination of the relevance of its disjuncts. For conjunction, however, we only had multiplication. But this latter can be changed when we take as our new relevance function the logarithm of $U(g, a)$.

Definition 9 (Reduction of surprisal) $V(g, a) = \log U(g, a)$.

This notion is well-known from Shannon's information theory (Cover & Thomas 1991): it measures the correlation between the propositions g and a. It is used to define the *mutual information* between two partitions. If G and A denote partitions of the same state space, the mutual information between them, $I(G, A)$, is defined as follows:

$$I(G, A) \quad = \quad \sum_{g \in G} \sum_{a \in A} P(g \cap a) \times V(g, a)$$

Just as $V(g, a)$ measures the correlation between the propositions g and a,[15] $I(G, A)$ measures the correlation between two partitions. It is a symmetric notion that is maximal in case $G = A$ and minimal in case the partitions are orthogonal to one another.

As we will see below, our new relevance function $V(\cdot, \cdot)$ shares a number of nice features with Good's $G(\cdot, \cdot)$. First of all, fact 35 immediately carries over to the new relevance relation.[16]

Fact 38 $V(a, g) = V(g, a)$.

With respect to disjunction, we have the following:

Fact 39 *If $P(a \cap b) = 0$, then $V(g, a \cup b) \leq \alpha V(g, a) + (1 - \alpha)V(g, b)$ for any $\alpha \in [0, 1]$, with equality if α is 1 or 0.*

We can obtain this result from fact 36 together with the fact that the function $\log x$ is strictly concave. The result then follows immediately from Jenning's equality (as explained in Cover & Thomas, 1991).

The new relevance function differs from Keynes' mainly through its behavior on *conjunctive* sentences. Using fact 37, one can easily show that in distinction to $U(g, a \cap b)$, the function $V(g, a \cap b)$ is – in natural special cases – additive with respect to the relevance of its conjuncts a and b.

Fact 40 *If $[(a \perp b)_P]$ and $[(a \perp b|\ g)_P]$, then $V(g, a \cap b) = V(g, a) + V(g, b)$.*

$$\begin{array}{llcll} & U(g, a \cap b) & = & U(g, a) \times U(g, b) & \text{if } [(a \perp b)_P] \ \& \ [(a \perp b|\ g)_P] \\ \text{iff} & \log U(g, a \cap b) & = & \log U(g, a) + \log U(g, b) & \text{if } [(a \perp b)_P] \ \& \ [(a \perp b|\ g)_P] \\ \text{iff} & V(g, a \cap b) & = & V(g, a) + V(g, b) & \text{if } [(a \perp b)_P] \ \& \ [(a \perp b|\ g)_P]. \end{array}$$

Notice that the requirements for relevance functions G and V to behave additively under conjunction differ: whereas G demands that $P(a \cap b/g) = P(a/g) \times P(b/g)$ and $P(a \cap b/\bar{g}) = P(a/\bar{g}) \times P(b/\bar{g})$, V demands that $P(a \cap b/g) = P(a/g) \times P(b/g)$ and $P(a \cap b/\top) = P(a/\top) \times P(b/\top)$.

[15] This is, of course, also the case for the commutative functions $C(\cdot, \cdot)$ and $U(\cdot, \cdot)$.

[16] Of course, in order for $V(g, a) \gtrless 0$, it has to be the case that $P(a) \neq 0$, $P(h) \neq 0$, $P(h/a) \neq 0$, and $P(a/h) \neq 0$, while for $V(g, \bar{a}) \gtrless 0$, it has to be the case that $P(\bar{a}) \neq 0$, $P(h) \neq 0$, $P(h/a) \neq 0$, and $P(\bar{a}/h) \neq 0$.

The title of this section suggests that the notion of relevance under discussion can be thought of as the reduction of the surprisal of a certain proposition. And indeed, this is so because if we define the surprisal of g given a, $\inf(g/a)$, by $\log \frac{1}{P(g/a)}$, we can prove the following fact.

Fact 41 *$V(g, a) = \inf(g) - \inf(g/a)$.*

$$\begin{array}{rcl} V(g,a) & = & \log U(g,a) \\ & = & \log \frac{P(g/a)}{P(g)} \\ & = & \log P(g/a) - \log P(g) \\ & = & -\log P(g) - -\log P(g/a) \\ & = & \inf(g) - \inf(g/a). \end{array}$$

Thus, this new notion of relevance measures the reduction of surprise, or of informativity, that g is true, due to the learning of a. Note that if a entails g, then $P(g/a) = 1$ and thus $\inf(g/a) = 0$. Thus, if $a \models g$, $V(g, a) = \inf(g)$.

Now suppose that proposition g expresses the goal-proposition to know what the extension is of a certain predicate C. The goal to learn what the world is like is in some sense more abstract than the goal that a certain particular proposition becomes true, if this particular proposition is not in itself indexically dependent on the actual world. In contrast to the proposition that John came to the party, for instance, the proposition that expresses who came to the party is indexically dependent on what the world is like, because in different worlds different individuals might have come to the party. This more abstract goal to know who came might actually be thought of as a *question*, and the uncertainty who will come can be measured by the entropy of the partition induced by the question, i.e., the entropy of $\{g_w | w \in W\} = \{\{v \in W | C(v) = C(w)\} | w \in W\}$. As we saw in section 2.2, the entropy of Q, $E(Q)$, is just the average informativity of the elements of partition Q: $\sum_{q \in Q} P(q) \times \inf(q)$. Now, the relevance of a according to our function $V(\cdot, \cdot)$ if the goal is to learn who has property C, i.e, which element of G is true, is just the reduction of entropy of G due to a.

Fact 42 *If G denotes a partition and the goal is to know which element of G is true, then $V(G, a)$ is the reduction of entropy of G due to a.*

$$\begin{array}{rcl} V(G,a) & = & \sum_{g \in G} P(g) \times V(g,a)] \\ & = & \sum_{g \in G} P(g) \times [\inf(g) - \inf(g/a)] \\ & = & [\sum_{g \in G} P(g) \times \inf(g)] - [\sum_{g \in G} P(g) \times \inf(g/a)] \\ & = & E(G) - E_a(G) = IV(G,a). \end{array}$$

Thus, we can define a natural relevance function that has many appealing features in common with the function used by Merin, but that is much closer related to a natural notion of relevance in case of pure information transmission than the latter notion.

10 Conclusion

In this paper I have discussed several possible candidates that can function as the 'default pragmatic value' of a sentence. All relevance functions defined relative to a particular goal-proposition had many features in common. Indeed, we saw that they all give rise to the same ordering relation between propositions. In this sense they are really different from $P(\cdot)$, $\inf(\cdot)$, and $IV(Q, \cdot)$. Although all the goal-oriented notions of relevance discussed give rise to the same ordering relations between propositions, for some the numerical values of complex sentences behave more 'linearly' than for others. Whether this is a decisive advantage or not depends crucially on what the pragmatic value is used for.

But even if two relevance functions behave very similarly under certain default assumptions, they sometimes give rise to contrasting predictions as to what these default assumptions are. If numerical values turn out to be really important for our understanding of natural language use, the search for the correct default assumptions become important. But that has to be left for another time.

References

[1] Adams, E. (1965), 'The logic of conditionals', *Inquiry*, **8**: 166-197.

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

[3] Bar-Hillel, Y. & R. Carnap (1953), 'Semantic information', *The British Journal for the Philosophy of Science*, **IV**, 147-157.

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

[5] Grice, H. P. (1967), 'Logic and Conversation', typescript from the William James Lectures, Harvard University. Published in P. Grice (1989), *Studies in the Way of Words*, Harvard University Press, Cambridge Massachusetts, pp. 22-40.

[6] Good, I.J. (1950), *Probability and the Weighing of Evidence*, Griffin, London.

[7] Hintikka, J. (1968), 'The varieties of information and scientific explanation', In: B. van Rootselaar & J.F. Staal (eds.), *Logic, Methodology and Philosophy of Science III*, North-Holland, Amsterdam, pp. 311-331.

[8] Horn, L. (1989), *A Natural History of Negation*, University of Chicago Press, Chicago.

[9] Jeffrey, R. (1965), *The Logic of Decision*, McGraw-Hill, New York.

[10] Keynes, J. M. (1921), *A Treatise on Probability*, Macmillan, London.

[11] Levinson, S. (2000), *Presumptive Meanings*, MIT Press, Cambridge MA.

[12] Lewis, D. (1976), 'Probabilities of conditionals and conditional probabilities', *Philosophical Review*, **85**: 297-315.

[13] Merin, A. (1993), *Sociomorph Linear Semantics*, draft, University of Munich.

[14] Merin, A. (1997), *If all our arguments had to be conclusive, there would be few of them*, Arbeitspapiere SFB 340, nr. 101, Stuttgart.

[15] Merin, A. (1999a), 'Information, relevance, and social decisionmaking: Some principles and results of decision-theoretic semantics', In; L.S. Moss, et al. (eds.), *Logic, Language, and Computation, Vol. 2*, CSLI Publications, Stanford, pp. 179-221.

[16] Merin, A. (1999b), *Die Relevanz der Relevanz: Fallstudie zur formalen Semantik der englishen Konjunktion 'but'*, Arbeitspapiere SFB 340, nr. 142, Stuttgart.

[17] Peirce, C.S. (1878), 'The probability of induction', *Popular Science Monthly*, **12**: 705-18.

[18] Rooij, R. van (2003), 'Quality and quantity of information exchange', *Journal of Logic, Language and Information*, **12**: 423-451.

[19] Rooij, R. van (2004), 'Utility, informativity and protocols', *Journal of Philosophical Logic*, **33**: 389-419.

[20] Sperber, D. and D. Wilson (1986), *Relevance*, Blackwell, Oxford.