On Quantifying Semantic Information

Simon D'Alfonso

November 3, 2010

Abstract

The purpose of this paper is to examine existing methods of semantic information quantification and suggest some alternatives. It begins with an outline of Bar-Hillel and Carnap's theory of semantic information before going on to look at Floridi's theory of strongly semantic information. The latter then serves to initiate an in-depth investigation into the idea of utilising the notion of truthlikeness to quantify semantic information. Firstly, a couple of approaches to measuring truthlikeness are drawn from the literature and explored, with a focus on their applicability to semantic information quantification. Secondly, a similar but new approach to measure truthlikeness/information is presented.

The term 'information' can mean many things. Here information is understood as factual semantic content, where if s is an instance of semantic content then:

1. s consists of one or more data
2. the data in s are well-formed
3. the well-formed data in s are meaningful

'Factual' simply means that the semantic content is about some state of affairs, that it is associated with a fact [1]. Ultimately semantic information is reducible to an identification with propositions (i.e. the information that p) and throughout this paper information is represented using statements in a basic propositional logic. Finally, as will become clearer, a veridicality thesis

[1] For more about all of this see [5]

for semantic information is endorsed, where if i counts as an instance of semantic information then it is semantic content that is also true [2].

So information is carried by statements/propositions and judgements such as 'statement A yields more information or is more informative than statement B' are naturally made [3]. Motivations for and aims of this enterprise to develop ways to quantify semantic information can be straightforwardly appreciated. The following cases and considerations will serve to illustrate this:

- Imagine a situation in which a six-sided die is rolled and it lands on 4. Given the following collection of statements describing the outcome of the roll, which is the most informative? Which is the least informative?
  - The die landed on 1
  - The die landed on 1 or 2 or 4
  - The die landed on 4
  - The die landed on 4 or 5
  - The die did not land on 4

  A formal framework for the quantification of semantic information could assign a numerical measure to each statement and rank accordingly.

- Take a simple domain of inquiry, involving the following two questions:
  - Is Berlin the capital city of Germany? (A)
  - Is Paris the capital city of France? (B)

  Now, the following ranking of answer statements in terms of highest to lowest informativeness seems to be right:

  1. $A \wedge B$
  2. $A \wedge \neg B$
  3. $\neg A \wedge B$
  4. $\neg A \wedge \neg B$

[2] See [3] and [4]

[3] Throughout this paper information yield is synonymous with informativeness

What is a suitable formal framework for the quantification of semantic information that can rigorously capture these intuitions and provide a way to measure the complete range of statements within a domain of inquiry?

- Take a more general example of a logical database consisting of a collection of facts. How can the information content of the database be measured? Why is one database more informative than another?

- Information is often treated as a commodity and one factor that will determine the value of a piece of information is going to be its quantitative measure. A formal method to measure the information yield of a statement could therefore facilitate its informational valuation.

Thus motivations and aims for the task of this paper are clear.

1 Bar-Hillel and Carnap's Theory of Semantic Information

Bar-Hillel and Carnap's seminal account of semantic information [1], henceforth referred to as Classical Semantic Information (CSI), measures the information yield of a statement within a given language in terms of the set of possible states it rules out and a logical probability space over those states. The general idea is based on the *Inverse Relationship Principle*, according to which the amount of information associated with a proposition is inversely proportional to the probability associated with that proposition. Using some *a priori* logical probability measure m on the space of possible states, two measures of information ($cont$ and inf) are provided, such that:

$$\mathrm{cont}(A) =_{df} 1 - m(A)$$

and

$$\mathrm{inf}(A) =_{df} -log_2(m(A))$$

In order to demonstrate these definitions, take a very simple propositional logical space consisting of three atoms.[4] The scenario concerns the canonical weather framework in the truthlikeness

[4]The spaces considered in this paper are finite

literature, where the three properties of a situation being considered are whether or not it will be (1) hot (h), (2) rainy (r) and (3) windy (w). Since there are three variable propositions involved, there are eight logically possible ways things can be, there are eight possible states. Here is a truth table depicting this:

State	h	r	w
w_1	T	T	T
w_2	T	T	F
w_3	T	F	T
w_4	T	F	F
w_5	F	T	T
w_6	F	T	F
w_7	F	F	T
w_8	F	F	F

Now, each possible state can be represented using a *state description*, a conjunction of atomic statements consisting of each atom in the logical space or its negation, but never both. For example, the state description for w_1 is $h \wedge r \wedge w$ and the state description for w_8 is $\neg h \wedge \neg r \wedge \neg w$. For the sake of example, the probability distribution used will be a simple uniform one; that is, since there are eight possible states, each state has a probability of $\frac{1}{8}$ that it will obtain or be actual.[5] Using these parameters, here are some results:

Statement A	conf(A)	inf(A)
$h \wedge w \wedge r$	0.875	3
$\neg h \wedge \neg w \wedge \neg r$	0.875	3
$h \wedge r$	0.75	2
$h \vee w$	0.25	0.415
$h \vee \neg h$	0	0
$h \wedge \neg h$	1	∞

Note that with the formulas for cont() and inf(), it is clear that the following hold between logical entailment and semantic information.

[5] Carnap discusses various types of more sophisticated logical probability measures. See http://plato.stanford.edu/entries/probability-interpret/#LogPro for an example.

$$A \vdash B \Rightarrow \mathrm{cont}(A) \geq \mathrm{cont}(B)$$

$$A \vdash B \Rightarrow \inf(A) \geq \inf(B)$$

Does the CSI account overall provide an acceptable measure of semantic information, in line with the motivations and aims discussed in the introduction? To a certain extent its results accord with our intuitions and are expected. True state descriptions yield much information; a statement which narrows things down to just one of the possible states yields a lot of information. By contrast, if a statement tells us that it will be hot and rainy but tells us nothing about whether or not it will be windy ($h \wedge r$), it would yield less information, as the calculations agree. Further still, a statement which only says that it will be hot or windy ($h \vee w$) yields relatively little information. Finally, a tautological statement, which gives no reduction of the original eight possibilities, appropriately yields no information. Despite these results, there are significant cons that indicate the inadequacy of the CSI account in accommodating certain criteria associated with the type of quantitative account of semantic information that is in line with present motivations and aims.

The most prominent issue concerns its assignment of maximal information yield to contradictions, what has elsewhere been termed the *Bar-Hillel-Carnap Paradox* [5]. Can no non-contradiction provide more information than a contradiction? Surely this is not the case. Furthermore, do contradictions yield any information at all? These questions will be discussed shortly.

Bar-Hillel and Carnap provide the following commentary on the matter:

> It might perhaps, at first, seem strange that a self-contradictory sentence, hence one which no ideal receiver would accept, is regarded as carrying with it the most inclusive information. It should, however, be emphasised that semantic information is here not meant as implying truth. A false sentence which happens to say much is thereby highly informative in our sense. Whether the information it carries is true or false, scientifically valuable or not, and so forth, does not concern us. A self-contradictory sentence asserts too much; it is too informative to be true. [1, pg. 229]

No doubt the notion of information expounded in the above passage is at odds with the ordinary sense of information. It is fair to say that the more a sentence says the more informative it is. But when does a sentence 'say much'? We intuitively judge the statement $A \wedge B$ to say more than the statement $A \vee B$ not just because it is less probable or excludes more possible states, but also because it does a better, more detailed job of describing how presumably things are. For any two true statements A and B such that $m(A) > m(B)$, it is fair to say that A yields more information than B. On the other hand, not only is $A \wedge \neg A$ false, but it does not at all do a good job of describing how things presumably are or could be. It does not discriminate and selectively narrow down on a potential state of affairs (unless a contradiction does actually occur!).

Further to this issue, the CSI account's indifference to truth and falsity means that it cannot distinguish between true and false statements with the same probability. If it actually is hot, rainy and windy (i.e. $h \wedge r \wedge w$ is the true state description), then in the sense of information we are interested in, the statement $h \wedge r \wedge w$ yields more information than the statement $\neg h \wedge \neg r \wedge \neg w$. Even the statement $h \wedge r$, which has a higher probability, yields more information than $\neg h \wedge \neg r \wedge \neg w$, since it contains more truth. These considerations suggest that at the least semantic information is not just about content and should be meant as implying truth.

2 Floridi's Theory of Strongly Semantic Information

In part prompted by these considerations, Luciano Floridi has developed a *Theory of Strongly Semantic Information* (TSSI), which differs fundamentally to CSI. It is termed 'strongly semantic' in contrast to Bar-Hillel and Carnap's 'weakly semantic' theory because unlike the latter, where truth values do not play a role, with the former semantic information encapsulates truth.

Floridi's basic idea is that the more accurately a statement corresponds to the way things actually are, the more information it yields. Thus information is tied in with the notion of *truthlikeness*. If a statement perfectly corresponds to the way things actually are, if it completely describes the truth of a domain of inquiry, then it yields the most information. Then there are two extremes. On the one hand, if a statement is necessarily true in virtue of it being a tautology, it yields no information. On the other, if a statement is necessarily false in virtue of it being a contradiction, it also yields no information. Between these two extremes and perfect correspondence there are contingently true statements and contingently false statements, which have varying degrees of information yield; the more truth over falsity the higher the information yield and the more falsity over truth the lower the information yield.

Using the weather framework introduced above, let the actual state be w_1 (throughout the remainder of this paper w_1 is assumed to be the actual state in all examples). The following ranking of propositions, from highest information yield to lowest information yield illustrates this idea:

1. $h \wedge r \wedge w$
2. $h \wedge r$
3. $\neg h \wedge \neg r$
4. $\neg h \wedge \neg r \wedge \neg w$
5. $h \wedge \neg h$, $h \vee \neg h$

It is time to take a look in more detail at Floridi's account as given in [2]. The model Floridi uses to illustrate his account, denoted E, has two predicates (G and H) and three objects (a, b and c). So in all there is a set of 64 possible states $W = w_1, ..., w_{64}$. Here are some key points:

- Each statement σ_i[6] in E has the property of providing true contents about (perfectly corresponding to) its corresponding situation w_i in W. $\Vdash_{w_i} \sigma_i$ is situation theoretic terminology indicating that w_i fully supports σ_i.
- the information yield of each σ_i is evaluated as a function of
  - the truth value of σ_i.
  - the degree θ of semantic deviation (degree of discrepancy) between σ_i and the actual situation w ($\Vdash_{w\theta} \sigma$).
- the degree of discrepancy for a statement σ from the actual situation is measured by a function f, which takes the statement as input and outputs some value in the interval [-1, 1]. This allows for the expression of both positive (when σ is true) and negative (when σ is false) degrees of discrepancy. Basically, the more a statement deviates from 0, the less information it yields.

[6]Floridi uses σ to stand for an instance of the situation theoretic infon. This distinction is of no concern for the purposes of this paper, so σ is treated as a statement.

Formally, Floridi stipulates the following five conditions, that "any feasible and satisfactory metric will have to satisfy" [2, p. 206.]:

(M.1) $\Vdash_W \sigma \rightarrow f(\sigma) = 0$

(M.2) $\Vdash_{\forall w} \sigma \rightarrow f(\sigma) = 1$

(M.3) $\Vdash_{\neg\exists w} \sigma \rightarrow f(\sigma) = -1$

(M.4) $\Vdash_{W-\theta} \sigma \rightarrow (0 > f(\sigma) > -1)$

(M.5) $\Vdash_{W+\theta} \sigma \rightarrow (0 < f(\sigma) < 1)$

The first condition covers cases where the statement is true and conforms to the actual situation W most precisely and accurately. The second condition covers cases where the statement being evaluated is a tautology, since all situations support it. The third condition covers cases where the situation is a contradiction, where no situations support it. The fourth condition deals with cases where the statement being evaluated is contingently false. For cases that fall under the fourth condition, f measures *degree of inaccuracy* and is calculated as the negative ratio between the number of false atomic statements (e) in the given statement σ and the length (l) of σ.

$$Inaccuracy : f(\sigma) = -\tfrac{e(\sigma)}{l(\sigma)}$$

To get values for e and l, it seems that some simplifying assumptions need to be made about a statement's form. Statements to be analysed in this way are in conjunctive normal form, with each conjunct being an atomic statement. e is the number of false conjuncts and l is the total number of conjuncts. This point will be discussed in more detail shortly.

The fifth condition deals with cases where the statement being evaluated is contingently true but does not conform to W with the highest degree of precision. For cases that fall under the fifth condition, f measures *degree of vacuity* and is calculated as the ratio between the number of situations, including the actual situation, with which σ is consistent (n) and the total number of possible situations (S) or states.

$$Vacuity : f(\sigma) = \tfrac{n(\sigma)}{S}$$

Of the inaccuracy metric Floridi writes, "[it] allows us to partition $\Sigma = s^l$ into l disjoint classes of inaccurate σ $\{\text{Inac}_1, ..., \text{Inac}_l\}$ and map each class to its corresponding degree of inaccuracy" [2, pg. 208]. An application to the model E is presented in the following table:[7]

# erroneous atomic messages in σ_i	Classes of inaccuracy	Cardinality of Inac_i	Degrees of inaccuracy	Degree of informativeness
1	Inac_1	6	$-\frac{1}{6}$	≈ 0.972
2	Inac_2	15	$-\frac{1}{3}$	≈ 0.888
3	Inac_3	20	$-\frac{1}{2}$	0.75
4	Inac_4	14	$-\frac{2}{3}$	≈ 0.555
5	Inac_5	7	$-\frac{5}{6}$	≈ 0.305
6	Inac_6	1	-1	0

For vacuous statements, Floridi writes "[the vacuity metric] and the previous method of *semantic weakening* allow us to partition Σ into $l-1$ disjoint classes $\text{Vac} = \{\text{Vac}_1, ..., \text{Vac}_{l-1}\}$, and map each class to its corresponding degree of vacuity" [2, pg. 209]. The semantic weakening method referred to consists of generating a set of statements by the following process. In this case, the number of atomic propositions in a statement is 6. Start with a statement consisting of 5 disjunctions, such as

$$Ga \vee Gb \vee Gc \vee Ha \vee Hb \vee Hc$$

This is the weakest type of statement and corresponds to Vac_1. Next, replace one of the disjunctions with a conjunction, to get something like

$$(Ga \vee Gb \vee Gc \vee Ha \vee Hb) \wedge Hc$$

This is the second weakest type of statement and corresponds to Vac_2. Continue this generation process until only one disjunction remains. The following table summarises an application of this process to the model E:

[7] Degree of informativeness calculations will be covered later

# compatible situations including w	Classes of vacuity	Cardinality of Vac_i	Degrees of vacuity	Quantity of informativeness
63	Vac_1	63	$\frac{63}{64}$	0.031
31	Vac_2	31	$\frac{31}{64}$	0.765
15	Vac_3	15	$\frac{15}{64}$	0.945
7	Vac_4	7	$\frac{7}{64}$	0.988
3	Vac_5	3	$\frac{3}{64}$	0.998

Say that the actual situation corresponds to the state description $Ga \wedge Ha \wedge Gb \wedge Hb \wedge Gc \wedge Hc$. Then the following tables summarises an example member of each class:

Class	Statement
Inac_1	$Ga \wedge Ha \wedge Gb \wedge Hb \wedge Gc \wedge \neg Hc$
Inac_2	$Ga \wedge Ha \wedge Gb \wedge Hb \wedge \neg Gc \wedge \neg Hc$
Inac_3	$Ga \wedge Ha \wedge Gb \wedge \neg Hb \wedge \neg Gc \wedge \neg Hc$
Inac_4	$Ga \wedge Ha \wedge \neg Gb \wedge \neg Hb \wedge \neg Gc \wedge \neg Hc$
Inac_5	$Ga \wedge \neg Ha \wedge \neg Gb \wedge \neg Hb \wedge \neg Gc \wedge \neg Hc$
Inac_6	$\neg Ga \wedge \neg Ha \wedge \neg Gb \wedge \neg Hb \wedge \neg Gc \wedge \neg Hc$
Vac_1	$Ga \vee Gb \vee Gc \vee Ha \vee Hb \vee Hc$
Vac_2	$(Ga \vee Gb \vee Gc \vee Ha \vee Hb) \wedge Hc$
Vac_3	$(Ga \vee Gb \vee Gc \vee Ha) \wedge Hb \wedge Hc$
Vac_4	$(Ga \vee Gb \vee Gc) \wedge Ha \wedge Hb \wedge Hc$
Vac_5	$(Ga \vee Gb) \wedge Gc \wedge Ha \wedge Hb \wedge Hc$

With a way to calculate degrees of vacuity and inaccuracy at hand, Floridi then provides a straightforward ways to calculate degrees of informativeness (g), by using the following formula, where once again f stands for the degree of vacuity/inaccuracy function:

$$g(\sigma) = 1 - f(\sigma)^2 \tag{1}$$

Furthermore, he also extends this and provides an extra way to measure amounts of semantic information. As this extension is simply derivative though and not essential, I will not go into it here. Suffice it to say, naturally, the higher the informativeness of σ the larger the quantity of semantic information it contains, and the lower the informativeness of σ the smaller the quantity of semantic information it contains. To calculate this quantity of semantic information contained in σ relative to $g(\sigma)$, Floridi makes use of integrals and the area delimited by the equation given for degree of informativeness.

2.1 Some Comments on Floridi's Theory

It will be evident to the reader that the classes of inaccuracy and vacuity presented by Floridi are not comprehensive in that they do not accommodate the full range of statements which could be constructed in the logical space of the model E. Once again, say that the actual situation corresponds to $Ga \wedge Ha \wedge Gb \wedge Hb \wedge Gc \wedge Hc$.

Take the false statement $Ga \wedge Ha \wedge Gb \wedge \neg Hb \wedge \neg Gc$, consisting of 5 conjoined atoms, 2 of which are false. A simple extension of Floridi's method for dealing with inaccuracy would naturally result in several other classes and the degree of inaccuracy of this statement would be $-\frac{2}{5}$.

Or take the following false statement:

$$\sigma = (Ga \vee Ha) \wedge Gb \wedge \neg Hb \wedge (\neg Gc \vee \neg Hc)$$

How should the formula given for inaccuracy be applied here? There is no clear-cut way to determine the values for e and l going by Floridi's description of the formula.

As for the possible classes of vacuity in E, it is clear that beyond those listed in above table, for any x such that $2 \leq x \leq 63$, it is possible to construct a statement such that the degree of vacuity is $\frac{x}{64}$.

In total the classes given by Floridi deal with 14 different propositions: 1 (true state description) + 1 (tautologies) + 1 (contradictions) + 6 (classes of inaccuracy) + 5 (classes of vacuity). Since there are 64 possible states in the space E, there are 2^{64} different propositions (propositions being satisfied by a set of states, so 2^{64} different sets of states).

With an aim of developing a system that can deal with the complete range of propositions in a given logical space, clearly Floridi's metric is inadequate.

Another aspect of this system which draws consideration is a certain asymmetry between the metrics for false statements and true statements; might there be one function that deals appropriately with both? A consequence of this separation between the false metric and true metric is that the same numerical value might be given to both a false statement and a true statement with otherwise different information measures. For example, take the following two statements:

1. $(A \wedge B) \vee (B \wedge C) \vee (C \wedge D)$ (true)
2. $A \wedge B \wedge \neg C \wedge \neg D$ (false)

Relative to an actual state corresponding to the state description $A \wedge B \wedge C \wedge D$, both 1 and 2 are given informativeness measures of 0.75. Yet it would seem that the first, true statement yields more information.

The central idea behind Floridi's theory is a right one. Both false and true statements can deviate from the precise truth. For false statements, the more falsity and less truth they contain, the greater the deviation hence the less information. For true statements, the less precise they are the greater their deviation, hence the less information. However as we have just seen it falls short of providing rigorous metrics, which can deliver appropriate and relatively consistent measures for the complete class of propositions in a logical space.

3 Information Quantification via Truthlikeness

As we have seen so far, the CSI approach to quantifying semantic information does not stipulate a veridicality condition for information. Floridi's TSSI on the other hand holds that semantic information should be meant as implying truth. This paves the way for an alternative approach to quantifying semantic information; rather than measuring information in terms of probability, the information given by a statement is to be measured in terms of its truthlikeness. It is indicated however that Floridi's contribution in [2] is of more conceptual rather than technical value, as the metrics provided can make way for some more detailed and refined improvements.

There already is a mass of formal work on truthlikeness to draw from.[8] Indeed the truthlikeness research enterprise has existed for at least a few decades now, far before any idea of utilising it to quantify semantic information. Whilst there are quite a few approaches to truthlikeness in the literature, to keep things simple and in some cases for certain technical reasons, I will look at two of them. The presentation given here of these two accounts is kept relatively simple and guided by the task at hand, which is to investigate their applicability to semantic information quantification. While seemingly a simple concept in essence, the notion of truthlikeness has resisted a straightforward formal characterisation. Over the last few decades of research, a variety of rival approaches have developed, each with their own pros and cons. In turn it follows that information quantification via truthlikeness measures is also not going to be a straightforward matter with a definitive account.

We start with a prominent and elegant approach to truthlikeness first proposed by Pavel Tichy and expanded upon by Graham Oddie [8, pg. 44]. This approach is an example of approaches that measure the truthlikeness of a statement A by firstly calculating its distance from a statement T using some distance metric Δ, where T is a state description of the actual state (so T is in a sense *the truth*). Actually, the distance metric is ultimately a function that operates on states. $\Delta(w_i, w_j)$ measures the distance between states w_i and w_j and we use the notation $\Delta(A, T)$ in effect as shorthand for an operation that reduces to Δ operation on states corresponding to A and T.

The result of this distance calculation is then used to calculate the statement's truthlikeness (Tr); the greater this distance, the less truthlike the statement and vice versa. This inverse proportionality is achieved simply with the following formula:

$$\mathrm{Tr}(A, T) = 1 - \Delta(A, T)$$

To see this approach at work, consider again the canonical weather framework, with w_1 being the actual state ($T = h \wedge r \wedge w$):

[8] Truthlikeness is also referred to as verisimilitude, although some use the two terms distinctively, to refer to a basic distinction in the types of approaches (verisimilitude for *content* approaches and truthlikeness for *likeness* approaches). Its origins can be traced back to Popper, who motivated by his philosophy of science, was the first philosopher to take the formal problem of truthlikeness seriously. See [9] for a brief introduction to truthlikeness. [8] and [6] are major and now somewhat classic pieces of work within the truthlikeness enterprise. A great piece of relatively recent work with a good summary of all that has come before it can be found in [13].

State	h	r	w
w_1	T	T	T
w_2	T	T	F
w_3	T	F	T
w_4	T	F	F
w_5	F	T	T
w_6	F	T	F
w_7	F	F	T
w_8	F	F	F

Before continuing with some examples, the introduction of some terminology is in order. We firstly recall that a *state description* is a conjunction of atomic statements consisting of each atom in the logical space or its negation, but never both and that each state description corresponds to a state. A statement is in *distributive normal form* if it consists of a disjunction of the state descriptions of states in which it is true. For example, $h \wedge r$ is true in states w_1 and w_2, so its distributive normal form is $(h \wedge r \wedge w) \vee (h \wedge r \wedge \neg w)$. For notational convenience throughout the remainder of this paper, when listing a statement in its distributive normal form, state descriptions may be substituted by the states they correspond to. For example, $h \wedge r$ may be represented with the term $w_1 \vee w_2$.

Now, take the possible states w_1, w_8 and w_5, which correspond to the state descriptions $h \wedge r \wedge w$, $\neg h \wedge \neg r \wedge \neg w$ and $\neg h \wedge r \wedge w$ respectively. The difference between w_1 and w_8 is {h, r, w}; they differ in every atomic assertion. w_5 and w_1 on the other hand differ by only {h}. So $\neg h \wedge r \wedge w$ is more truthlike than $\neg h \wedge \neg r \wedge \neg w$, because the distance between w_5 and the actual state w_1 is less than the distance between w_8 and w_1.

The general formula to calculate the distance between A and T is:

$$\Delta(A, T) = \frac{d}{|W_A|} \tag{2}$$

where

- Let w_T stand for the state that corresponds to T.

- W_A stands for the set of states in which A is true.

- A weight with value $\frac{1}{n}$ is assigned to every atomic state, where n is the number of propositions in the logical space. This weight is used to calculate the value of the distance between two states, with each atomic difference adding $\frac{1}{n}$. So $\Delta(w_5, w_1) = \frac{1}{3}$ and $\Delta(w_8, w_1) = 1$.

- d is the sum of atomic assertion differences between A and T. That is, the sum of $\Delta(w_a, w_T)$ for each $w_a \in W_A$.

So the statement $\neg h \wedge r \wedge w$ (the state description for w_5) has a truthlikeness (information measurement for our purposes) of $\frac{2}{3}$ and the statement $\neg h \wedge \neg r \wedge \neg w$ (the state description for w_8) has a truthlikeness (information measurement) of 0.

This approach extends to the complete range of statements involving h, r and w. According to Formula 2, the distance of a statement from the truth is defined as the average distance between each of the states in which the state is true and the actual state. Take the statement $h \wedge \neg r$, which makes assertions about only 2 of the 3 atomic states. It is true in both w_3 and w_4, or $\{h, \neg r, w\}$ and $\{h, \neg r, \neg w\}$ respectively. w_3 has a distance of 0.33 and w_4 has a distance of 0.67 from w_1 so the average distance is $\frac{0.33+0.67}{2} = 0.5$. Note that from henceforth the truthlikeness function (Tr) will be replaced by an information yield function (info), so that

$$\text{info}(A, T) = 1 - \Delta(A, T)$$

Also, given that T is set as w_1, $\text{info}(A, T)$ will generally be abbreviated to $\text{info}(A)$.

Table 1 lists some results using this method. How do these results fare as measures of information yield? 1 clearly yields the most information and 21 clearly yields the least. 10, 11 and 15 indicate that for a true disjunctive statement, the more false constituents it has the less information it yields. In general, the more false constituents contained in a formula the more likely a decrease in information yield, as indicated by the difference between 5 and 9.

Statements 7, 17 and 21 make sense; the greater the number of false conjuncts a statement has, the less information it yields, with the upper bound being a statement consisting of nothing but false conjuncts, which is accordingly assigned a measure of 0. Although 17 and 19 have the same measure, 17 has one true conjunct out of three atomic conjuncts and 19 has zero true atomic statements out of one atomic statement. From this it can be said that the assertion of a false statement detracts more from information yield than the absence of an assertion or denial of that statement. Half of 14 is true, so it is appropriately assigned a measure of 0.5. 20 further shows that falsity proportionally detracts from information yield. The difference between 18 and 16 is perhaps the most interesting. Although 18 has one true conjunct out of two and 16 contains no true atoms, 16 comes out as yielding more information, further suggesting the price paid for asserting falsity in a conjunction. Also, note that some false statements yield more information than some true statements.

One possible issue with this method is that tautologies are not assigned a maximum distance and hence are assigned a non-zero, positive measure. Without going into detail, it seems that this is a more significant issue for a quantitative account of information than it is for an account of truthlikeness. The implication that tautologies have a middle degree of information yield strongly conflicts with the intuition and widely accepted position that tautologies are not informative. Whilst an explanation of this result will be discussed subsequently, the simplest and perhaps best way to get around this issue would be to just exclude tautologies from this metric and assign them a predefined distance of 1 hence information yield of 0. Although an expedient, this move would put the tautology alongside its extreme counterpart the contradiction, which is also excluded. In the case of contradictions (22), these calculations do not apply, because since they are true in no states this would mean division by zero.

3.1 Quantifying Misinformation

One task that Floridi leaves for a subsequent stage of research in his paper is "the extension of the quantitative analysis to the semantic concepts of quantity of misinformation" [2, p. 217.]. How might this approach to quantifying information be extended in order to quantify misinformation?

To begin with, it seems that one straightforward stipulation to make is that true statements should have a measure of 0, for they are not misinformative. So unlike information measures, where both true and false statements can have relevant measures, with misinformation attention is confined to false statements.[9] Of course, the more falsity in a false statement, the more misinformation it yields. As with the information measure just given, the first metric to devise is one which

[9]There is a sense in which true statements might contain misinformation, as will be discussed later.

#	Statement (A)	T/F	info(A)
1	$h \wedge r \wedge w$	T	1
2	$h \wedge r$	T	0.83
3	$h \wedge (r \vee w)$	T	0.78
4	$h \wedge (\neg r \vee w)$	T	0.75
5	$(h \wedge r) \vee w$	T	0.67
6	h	T	0.67
7	$h \wedge r \wedge \neg w$	F	0.67
8	$h \vee r$	T	0.61
9	$(h \wedge \neg r) \vee w$	T	0.6
10	$h \vee r \vee w$	T	0.57
11	$h \vee r \vee \neg w$	T	0.52
12	$h \vee \neg r$	T	0.5
13	$h \vee \neg h$	T	0.5
14	$h \wedge \neg r$	F	0.5
15	$h \vee \neg r \vee \neg w$	T	0.48
16	$\neg h \vee \neg r \vee \neg w$	F	0.43
17	$h \wedge \neg r \wedge \neg w$	F	0.33
18	$(h \vee \neg w) \wedge \neg r$	F	0.33
19	$\neg h$	F	0.33
20	$\neg h \wedge \neg r$	F	0.17
21	$\neg h \wedge \neg r \wedge \neg w$	F	0
22	$h \wedge \neg h$	F	N/A

Table 1: Information yield results using Tichy/Oddie metric

measures the degree of deviation from complete misinformation ($\Delta_{misinfo}$). A deviation of 0 will translate to maximum misinformation and a deviation of 1 will translate to no misinformation. So the metric for this deviation will at least satisfy these conditions:

1. all true statements have a predefined deviation of 1

2. all contingently false statements have a deviation greater than or equal to 0 and less than 1

Two ways to go about measuring this deviation come to mind. The first is to use the complement of the deviation for information measures: $\Delta_{misinfo}(A,T) = 1 - \Delta_{info}(A,T)$.

For the second, I'll use some terminology given by Oddie [8, p. ???.], who discusses a reversal operation Rew() on states such that:

Rew(U) = the state V such that for any atomic state B, B is true in U if and only if B is false in V.

This reversal operation on states is extended to a reversal operation on propositions. The reversal of a proposition A is the image of A under Rew():

$$\text{Rev}(A) = \text{the proposition } B \text{ such that } A \text{ contains state } U \text{ if and only if } B \text{ contains Rew}(U)\text{, for any state } U.$$

Where w_T is the actual world, the second way to measure misinformation deviation would be to measure the distance of a statement from Rew(w_T). In our example, $w_T = w_1$ so Rew$(w_T) = w_8$.

The the greater the distance, the less misinformation yielded by the statement. These two approaches turn out in fact to be equivalent. From here one can go on to calculate a statement's quantity of misinformation simply by subtracting its deviation or distance from 1.

$$\text{misinfo}(A) = 1 - \Delta_{misinfo}(A, T)$$

From this it follows that

$$\text{info}(A) + \text{misinfo}(A) = 1$$

and the following hold:

- $\text{info}(\text{Rev}(T)) = 0$
- $\text{info}(A) + \text{info}(\text{Rev}(A)) = 1$
- $\text{misinfo}(A) = \text{info}(\text{Rev}(A))$

As this section has shown, unlike the CSI framework, truthlikeness approaches such as this one accommodate semantic misinformation.

3.2 Adjusting Atom Weights

It is worth briefly mentioning the possibility of adjusting atomic weights in order to reflect differences in the informational value of atomic statements. As we have seen, where n stands for the number of propositional atoms in a logical space, each atom is assigned a standard weight of $\frac{1}{n}$ for the purposes of Δ calculation. In the 3-proposition weather example being discussed, this results in each atom being assigned a weight of $\frac{1}{3}$. As a consequence of this even distribution of weighting, the three statements $h \wedge r \wedge \neg w$, $h \wedge \neg r \wedge w$ and $\neg h \wedge r \wedge w$ are all assigned the same information yield measure.

Beyond this there is the possibility of adjusting the weightings so that the resulting assignment is non-uniform. Such a modification could perhaps be used to model cases where different atomic statements (hence respective compound statements containing them) have different informational value. There is much room to interpret just what is meant here by 'informational value'. Statement A could have a greater informational value than B if an agent prefers the acquisition of A over the acquisition of B, or if the agent can do more with A than B. Variable weightings could also perhaps be used to reflect extra-quantitative or qualitative factors. These are all just some offhand thoughts.

The minimum requirement is that the sum of the values assigned to the atoms comes to 1. With this in mind, take the following weighting assignments:

- h - $\frac{1}{6}$
- r - $\frac{2}{6}$
- w - $\frac{3}{6}$

With such an assignment, the information yield measurement distribution changes significantly. Some results are given in Table 2.

#	Statement (A)	info(A)
1	$h \wedge r \wedge \neg w$	0.5
2	$h \wedge \neg r \wedge w$	0.67
3	$\neg h \wedge r \wedge w$	0.83
4	$h \wedge \neg r \wedge \neg w$	0.167
5	$\neg h \wedge \neg r \wedge w$	0.5
6	$\neg h \wedge r \wedge \neg w$	0.33

Table 2: Results with adjusted weightings

It can be seen that although statements 1, 2 and 3 all share the same form (2 true atoms and 1 false atom), none share the same information yield measure. In such a case, 3 yields more information than 1 due to the fact that its two true atoms are of greater informational value than the two true atoms contained in 1.

One issue with such a modification is the treatment of misinformation quantification. Although 1 has a lower information yield measure than 3, it doesn't seem right to say that conversely it has a greater misinformation yield, since it contains the same amount of falsity as 3.

3.3 Contradictions

As we have just seen, it is not mathematically possible to calculate a value for contradictions using this truthlikeness method (since contradictions contain no states, this mean a division by 0). How then should contradictions be dealt with? One option is to simply exclude contradictions from the metrics and assign them a predefined deviation of 1 hence information yield of 0. As we have seen, this is the approach that Floridi takes.

This however is arguably too rash a move and there are good reasons to adopt an approach in which the metrics are adjusted or expanded in order to accommodate contradictions. To begin with, the class of contradictions is not homogeneous with regards to information yield and different contradictions can be treated as having different non-zero information yields.

Take a logical space consisting of 100 atomic propositions, $p_1, p_2 ... p_{100}$, all true relative to the actual state. Now the statement $p_1 \wedge p_2 \wedge p_3 \wedge ... \wedge p_{99} \wedge p_{100}$ yields maximal information. If we conjoin it with $\neg p_{100}$ to get the contradiction $p_1 \wedge p_2 \wedge p_3 \wedge ... \wedge p_{99} \wedge p_{100} \wedge \neg p_{100}$, the original information remains and the resulting statement should not instantly be assigned an information measure of 0. To put it another way, if one contradictory atom is inserted into a database with much information, whilst this means that there is now some misinformation within the database, surely there is still a lot of information within the database. If the contents of the database were to be represented as a statement (the statement being a contradiction), it would be good to have a way to measure information that can deal with the fact that the database still contains much information and that is also sensitive to the fact that different contradictions have different information yield measures; for example, $p_1 \wedge p_2 \wedge p_3 \wedge ... \wedge p_{99} \wedge p_{100} \wedge \neg p_{100}$ clearly yields more information than $p_1 \wedge \neg p_1$.

Now that the case has been made, in order to accommodate contradictions and have a way to assign them positive, non-zero measures of information, the following framework is suggested. All non-contradictions are dealt with as before, using the standard metrics with classical propositional logic. For contradictions, a many-valued framework such as that associated with the paraconsistent logic *LP* [10] is introduced.[11] In this framework, a third truth value, B (true and false) is introduced. Importantly, since contradictions can be true in states involving B, the denominator in calculations involving contradictions need no longer be 0.

In the classical framework, each atomic element is assigned a weight given by $\frac{1}{n}$ and the value of the distance between a T and an F is equal to this weight. For this extended framework, the classical distances are still the same and the distance between B and a T or an F is $\frac{1}{2n}$. Semantically speaking, the reason for this is that B consists of both T and F, so it half corresponds to either one of them. Formally, the distance calculation between A and T now becomes:

$$\Delta(A, T) = \frac{d + d_{\mathrm{B}}}{|W_A|} \tag{3}$$

where

- Let w_T stand for the state that corresponds to T.
- W_A stands for the set of states in which A is true.
- where n is the number of propositions in the logical space:
  - A weight with value $\frac{1}{n}$ is assigned to every classical atomic state (T or F). The distance function $\Delta(w_a, w_T)$ deals with calculations involving classical states.
  - A weight with value $\frac{1}{2n}$ is assigned to every non-classical atomic state (B). The distance function $\Delta_{\mathrm{B}}(w_a, w_T)$ deals with calculations involving the non-classical states.
- d is the sum of atomic assertion differences between A and T involving the truth values T and F. That is, the sum of $\Delta(w_a, w_T)$, for each $w_a \in W_A$.
- d_{B} is the sum of atomic assertion differences between A and T involving the truth value B. That is, the sum of $\Delta_{\mathrm{B}}(w_a, w_T)$, for each $w_a \in W_A$.

[10] http://plato.stanford.edu/entries/logic-paraconsistent/#ManyValLog

[11] The adoption of such a framework is for instrumental purposes only and is not an endorsement of paraconsistency or dialetheism.

To see this approach at work we again consider the weather framework. The list of 27 possible states looks like this:

	h	r	w		h	r	w		h	r	w
w_1	T	T	T	w_{10}	B	T	T	w_{19}	F	T	T
w_2	T	T	B	w_{11}	B	T	B	w_{20}	F	T	B
w_3	T	T	F	w_{12}	B	T	F	w_{21}	F	T	F
w_4	T	B	T	w_{13}	B	B	T	w_{22}	F	B	T
w_5	T	B	B	w_{14}	B	B	B	w_{23}	F	B	B
w_6	T	B	F	w_{15}	B	B	F	w_{24}	F	B	F
w_7	T	F	T	w_{16}	B	F	T	w_{25}	F	F	T
w_8	T	F	B	w_{17}	B	F	B	w_{26}	F	F	B
w_9	T	F	F	w_{18}	B	F	F	w_{27}	F	F	F

Take the statement $h \wedge \neg h$. It holds in $w_{10} - w_{18}$. In this set of states there are 6 instances of F and 15 instances of B. Therefore, the distance is:

$$\Delta(h \wedge \neg h, h \wedge r \wedge w) = \frac{(6 \times \frac{1}{3}) + (15 \times \frac{1}{6})}{9} = 0.5$$

Table 3 lists some results, with all statements being classical contradictions.

#	Statement (A)	info(A)
1	$h \wedge \neg h \wedge r \wedge w$	0.67
2	$h \wedge \neg h \wedge r \wedge \neg r \wedge w$	0.583
3	$((h \wedge \neg h) \vee (r \wedge \neg r)) \wedge w$	0.583
4	$h \wedge \neg h$	0.5
5	$(h \wedge \neg h) \vee (r \wedge \neg r)$	0.5
6	$(h \wedge \neg h) \vee (r \wedge \neg r \wedge w)$	0.5
7	$(h \wedge \neg h) \vee (r \wedge \neg r) \vee (w \wedge \neg w)$	0.5
8	$h \wedge \neg h \wedge \neg r$	0.417
9	$h \wedge \neg h \wedge \neg r \wedge \neg w$	0.33

Table 3: Results for contradictions

3.4 Truthlikeness Adequacy Conditions and Information Conditions

It was seen with the CSI account that if $A \vdash B$ then cont(A) $\geq$ cont(B) and inf(A) $\geq$ inf(B). This is not going to hold in general for a Tichy/Oddie truthlikeness approach to information quantification.

For example, looking back at Table 1, although $\neg h \wedge \neg r \wedge \neg w \vdash \neg h$, $\text{info}(\neg h \wedge \neg r \wedge \neg w) = 0 < \text{info}(\neg h) = 0.334$. However since both of these statements are false, this example is not an issue; given two false statements such as these, the logically stronger one is understandably further away from the truth.

The question of interest is whether or not this property holds when A and B are both instances of information, when they are both true. So the pertinent question is this: among true statements, does information as truthlikeness here covary with logical strength? Put formally, does the following condition hold?:

$$\text{If } A \text{ and } B \text{ are true statements and } A \vdash B, \text{ then } \text{info}(B) \leq \text{info}(A)$$

As it turns out, information/truthlikeness does not covary with logical strength using the Tichy/Oddie approach. To begin with, it can be seen that although $h \vee \neg r \vee \neg w \vdash h \vee \neg h$, $\text{info}(h \vee \neg r \vee \neg w) = 0.48$ whereas $\text{info}(h \vee \neg h) = 0.5$. But leaving aside cases where tautologies are involved, this result also does not hold more generally, in cases where only contingently true statements are involved. For example, although $(h \wedge \neg r) \vee w \vdash h \vee w$, $\text{info}((h \wedge \neg r) \vee w) = 0.6 < \text{info}(h \vee w) = 0.61$. This is not an isolated case either.

Remark In a logical space containing three propositional variables, there are 2187 possible ways to have two true statements A and B such that $A \vdash B$. Out of these, 366 are such that $\text{info}(A) < \text{info}(B)$.

An interesting example arises in a space with four atoms. Take the 4-proposition logical space obtained by adding the propositional variable d to the 3-atom weather framework, with the actual state corresponding to the state description $h \wedge r \wedge w \wedge d$. Whilst it is the case that $\neg h \vee \neg r \vee w \vdash \neg h \vee \neg r \vee w \vee \neg d$, $\text{info}(\neg h \vee \neg r \vee w) = 0.482$ whilst $\text{info}(\neg h \vee \neg r \vee w \vee \neg d) = 0.483$

This result seems quite counter-intuitive; the addition of a false disjunct to an already true statement slightly increases its information yield. Contrary to this result, it is fair to say that information is proportional to accuracy, and that the addition of disjunctions, particularly when false, decreases accuracy. Also, this result seems contrary to the notion of information loss. If the information $\neg h \vee \neg r \vee w$ stored in a database were corrupted and the salvaged remains were the weakened $\neg h \vee \neg r \vee w \vee \neg d$, it would ordinarily be said that say that this would be a case of information loss. Or if a signal is transmitting the message $\neg h \vee \neg r \vee w$ and it became $\neg h \vee \neg r \vee w \vee \neg d$ due to noise, once again it can be said that there is information loss.

The failure of Tichy/Oddie truthlikeness, hence for our purposes information, to covary with logical strength amongst true statements is discussed by Ilkka Niiniluoto, who whilst developing his own account of truthlikeness in [6] surveys how various approaches to truthlikeness fare against a range of adequacy conditions.[12] Though an interpretation of the Tichy/Oddie method that can serve to explain its failure to satisfy this condition and justify its use will be discussed shortly, in the meantime we will take a brief look at Niiniluoto's preferred approach to truthlikeness, which satisfies all of the adequacy conditions he lists.

3.5 Niiniluoto on Truthlikeness

As with the Tichy/Oddie approach to truthlikeness, the task for Niiniluoto is to define some distance function Δ such that $\Delta(A, T) \in [0, 1]$. He gives six distance functions, which are listed below. Firstly, a recap and establishment of terms to be used:

- $\Delta(w_i, w_j)$ calculates the distance between states w_i and w_j. This is the sum of atomic differences multiplied by the atomic weight ($\frac{1}{n}$)
- w_T is the actual state, that corresponds to the true state description T.
- W_A is the set of states in which A is true.
- $\mathbf{B}$ is the set of all states in the logical space.

Here are the distance functions:

- $\Delta_{\text{min}}(A, T)$ = the minimum of the distances $\Delta(w_a, w_T)$ with $w_a \in W_A$.
- $\Delta_{\text{max}}(A, T)$ = the maximum of the distances $\Delta(w_a, w_T)$ with $w_a \in W_A$.
- $\Delta_{\text{sum}}(A, T)$ = the sum of all distances $\Delta(w_a, w_T)$ with $w_a \in A$, divided by the sum of all distances $\Delta(w_b, w_T)$ with $w_b \in \mathbf{B}$.

[12] These conditions will be looked at in Section 4.1.

- $\Delta_{\text{av}}(A,T)$ = the sum of the distances $\Delta(w_a, w_T)$ with $w_a \in A$ divided by $|W_A|$.
- $\Delta^{\gamma}_{\text{mm}}(A,T) = \gamma\Delta_{\text{min}}(A,T) + (1-\gamma)\Delta_{\text{max}}(A,T)$ for some weight γ with $0 \leq \gamma \leq 1$
- $\Delta^{\gamma\lambda}_{\text{ms}}(A,T) = \gamma\Delta_{\text{min}}(A,T) + \lambda\Delta_{\text{sum}}(A,T)$ for some two weights γ and λ with $0 \leq \gamma \leq 1$ and $0 \leq \lambda \leq 1$

Δ_{av} is the Tichy/Oddie approach. Niiniluoto's preferred metric for truthlikeness is $\Delta^{\gamma\lambda}_{\text{ms}}$, which he terms the weighted min-sum measure [6, p. 228.]. This distance calculation is then used to calculate truthlikeness: $\text{Tr}(A,T) = 1 - \Delta^{\gamma\lambda}_{\text{ms}}(A,T)$. Table 4 lists some results using the min-sum measure, with γ being assigned the value 0.89 and λ being assigned the value 0.44.

#	Statement (A)	T/F	info(A)
1	$h \wedge r \wedge w$	T	1
2	$h \wedge r$	T	0.96
3	$h \wedge (r \vee w)$	T	0.93
4	$h \wedge (\neg r \vee w)$	T	0.89
5	h	T	0.85
6	$(h \wedge r) \vee w$	T	0.81
7	$(h \wedge \neg r) \vee w$	T	0.78
8	$h \vee r$	T	0.74
9	$h \wedge r \wedge \neg w$	F	0.67
10	$h \vee r \vee w$	T	0.67
11	$h \vee \neg r$	T	0.67
12	$h \vee r \vee \neg w$	T	0.63
13	$h \vee \neg r \vee \neg w$	T	0.6
14	$h \wedge \neg r$	F	0.59
15	$h \vee \neg h$	T	0.56
16	$(h \vee \neg w) \wedge \neg r$	F	0.48
17	$\neg h$	F	0.41
18	$h \wedge \neg r \wedge \neg w$	F	0.33
19	$\neg h \vee \neg r \vee \neg w$	F	0.26
20	$\neg h \wedge \neg r$	F	0.22
21	$\neg h \wedge \neg r \wedge \neg w$	F	0
22	$h \wedge \neg h$	F	N/A

Table 4: Information yield results using Niiniluto's min-sum measure

3.6 An interpretation of the Tichy/Oddie approach

Returning to the Tichy/Oddie approach, it is time to investigate further some of its problematic aspects which we touched upon earlier and see what we might make of them in relation to task of quantifying information. But before doing so, an initial general point to make is that when it comes to considerations of information/misinformation yield, there is a certain asymmetry between true and false statements. Whilst false statements (misinformation) can be judged to yield some truth (information), true statements are ordinarily judged to just contain information and do not yield any misinformation. But like false statements can yield some information, there is a sense in which true statements can yield some misinformation. Given an actual state corresponding to the state description $h \wedge r \wedge w$, it is straightforward to say that the false statement $h \wedge \neg r \wedge \neg w$ still yields some information. On the other hand, the statement $h \vee \neg w \vee \neg r$, whilst true, does not give one the complete truth about the domain of inquiry and the majority of its disjuncts are false. As will be shown now, such a statement can be seen as being potentially misinformative in the sense that it can potentially lead to falsity and it is this type of view that can be associated with and justify usage of the Tichy/Oddie approach for certain applications.

In Section 3.4 we saw that whilst it is the case that $(h \wedge \neg r) \vee w \vdash h \vee w$, $\text{info}((h \wedge \neg r) \vee w) < \text{info}(h \vee w)$. In light of the interpretation to be now provided, this result finds some support. $h \vee w$ contains no false atoms, whereas $(h \wedge \neg r) \vee w$ contains one, namely $\neg r$. Expanding upon this example to emphasise the point, take a logical space consisting of five atoms ($\{p_n | 1 \leq n \leq 5\}$), such that each is true in the actual state. Whilst it is the case that $(p_1 \wedge \neg p_2 \wedge \neg p_3 \wedge \neg p_4) \vee p_5 \vdash p_1 \vee p_5$, the antecedent statement makes many false assertions whereas the consequent makes none. In this way the Tichy/Oddie approach can be seen to measure not only how much atomic truth a statement contains, but also how much atomic falsity it contains. Also, in light of Section 3.1, these types of results can be seen as part of adopting a uniform method for the measurement of information and its complementary misinformation.

Something along the lines of the idea just outlined can be rigorously captured by viewing the Tichy/Oddie approach in terms of *expected utility*. We begin by assigning each state description a utility value, where the value for a state description S ($\text{val}(S)$) is determined using the following method:

- Let n stand for the number of propositional variables in the logical space
- Let t stand for the number of true atoms, relative to the actual state, in the state w corresponding to S
- $\text{val}(S) = \frac{t}{n}$

So in the case of our 3-atom logical space with w_1 being the actual state, each state description (represented using the state it corresponds to) is valued as follows:

- $\text{val}(w_1) = 1$
- $\text{val}(w_2) = \text{val}(w_3) = \text{val}(w_5) = \frac{2}{3}$
- $\text{val}(w_4) = \text{val}(w_6) = \text{val}(w_7) = \frac{1}{3}$
- $\text{val}(w_8) = 0$

Now, given a statement A that holds in n states, convert it to distributive normal form, which will have n disjuncts. Imagine an agent is to choose one and only one of the disjuncts. The value of the selected disjunct determines the utility or informational value of the choice. Using the standard decision theoretic framework, we can say that the estimated utility of A ($\text{eu}(A)$) is:

$$\text{eu}(A) = \sum_{i=1}^{n} \text{val}(S_i) \times \text{pr}(S_i)$$

where $\text{val}(S_i)$ is the value of a state description and $\text{pr}(S_i)$ is the probability of it being chosen. Each disjunct has the same probability of being selected as any other, so the probability of a disjunct being chosen is $\frac{1}{n}$. This estimated utility value equates to the statement's truthlikeness/information measure using the Tichy/Oddie approach.

Going back to the sound argument $(h \wedge \neg r) \vee w \vdash h \vee w$, here are both antecedent and consequent in distributive normal form followed by a decision-theoretic-style tabulation of the two statements:

- $(h \wedge \neg r) \vee w \equiv w_1 \vee w_3 \vee w_4 \vee w_5 \vee w_7$
- $h \vee w \equiv w_1 \vee w_2 \vee w_3 \vee w_4 \vee w_5 \vee w_7$

	w_1	w_3	w_4	w_5	w_7
$(h \wedge \neg r) \vee w$	1	$\frac{2}{3}$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{1}{3}$

$\text{eu}((h \wedge \neg r) \vee w) = (1 \times \frac{1}{5}) + (\frac{2}{3} \times \frac{1}{5}) + (\frac{1}{3} \times \frac{1}{5}) + +(\frac{2}{3} \times \frac{1}{5}) + (\frac{1}{3} \times \frac{1}{5}) = 0.6$

	w_1	w_2	w_3	w_4	w_5	w_7
$h \vee w$	1	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{1}{3}$

$$\mathrm{eu}(h \vee w) = (1 \times \tfrac{1}{6}) + (\tfrac{2}{3} \times \tfrac{1}{6}) + (\tfrac{2}{3} \times \tfrac{1}{6}) + (\tfrac{1}{3} \times \tfrac{1}{6}) + (\tfrac{2}{3} \times \tfrac{1}{6}) + +(\tfrac{1}{3} \times \tfrac{1}{6}) = 0.61$$

Thus the information measurement of a statement here can be seen in terms of its expected utility, whereby both the positive (truth, information) and the negative (falsity, misinformation) are factored into calculations. Analysing the Tichy/Oddie measure in this way explains its counter-intuitive results. Take the example in Section 3.4, where $\mathrm{info}(\neg h \vee \neg r \vee w) < \mathrm{info}(\neg h \vee \neg r \vee w \vee \neg d)$. Whilst the addition of a false disjunct to a statement is grounds to think that its information measure should decrease, seen in terms of expected utility, the addition of one false disjunct results in the addition of one more model to the set of satisfying models, namely the model corresponding to the state description $h \wedge r \wedge \neg w \wedge \neg d$. As can be seen, this addition actually results in a greater overall expected utility.

Apparently Tichy himself offered a similar analysis as an argument in defence of his truthlikeness measure [6, p. 238]. Niiniluoto is dismissive of the argument, but whilst his reasons are legitimate, they pertain specifically to considerations of truthlikeness. At this point it is important to stress that although we are using the notion of truthlikeness to quantify information, this is not to say that they amount to one and the same thing and a consideration in accounting for one need not be a consideration in accounting for the other. Furthermore, the concept of information is generally treated more pluralistically than truthlikeness, "it is hardly to be expected that a single concept of information would satisfactorily account for the numerous possible applications of this general field" [11]. This point of view could accomodate the suggestion that for some purposes of semantic information quantification this approach might be suitable. In fact, some of Niiniluoto's claims could be seen as offering support for the use of the Tichy/Oddie approach in a framework for information measurement. For example, he commences his case with the following:

> even if it were the case that function M_{av} [the Tichy/Oddie measure] serves to measure the degree of *trustworthiness* (or pragmatic preference for action) of a hypothesis, it would not follow that M_{av} is an adequate measure of *truthlikeness*. ... This idea is irrelevant to, if we are dealing with answers to cognitive problems - the connection between truthlikeness and practical action is quite another question which has to be studied separately.[6, p. 238]

Also, he concludes with:

> I conclude that Tichy's and Oddie's average measure for the logical concepts of truthlikeness rests upon intuitions which pertain to the epistemic problem of truthlikeness.[6, p. 241]

In closing and without elaboration, it suffices to mention that terms employed here, such as 'trustworthiness' and 'practical action' could be not inappropriately associated with the notion of information.

3.6.1 Adjusting State Description Utility Values

Given this decision-theoretic interpretation of the Tichy/Oddie method, one possible extension is to adjust the way in which utilities are assigned to state descriptions. As we have seen, the standard method uses a simple, linear utility function; for all levels l, there is a constant difference between l and $l+1$ / $l-1$. As a simple example of this extension, take a basic quadratic utility function. Where x stands for the standard linear utility of a state description, its modified utility value y is calculated as follows: $y = x^2$.

What does a utility function such as this one say in terms of state description distribution and information yield? Roughly speaking, the more efficiently one statement gets to containing a high valued state description disjunct the better, where more efficiently means doing so with less state description disjuncts.

In the ranking of all possible 255 statements in the 3-atom logical space, from highest to lowest information yield, the original method using a standard linear utility and the method with this quadratic utility function agree on the first eight positions, after which differences start to appear. To illustrate the higher premium given to relatively small state collections involving w_1, whilst the linear way places w_1, w_8 below w_1, w_2, w_8, the quadratic way places w_1, w_8 above w_1, w_2, w_8.

If the rate of increase of the utility function is sufficiently large, it is even possible to get the covariation with logical strength amongst true statements condition to hold, although at the cost of a skew where a disproportionate number of states containing w_1 are placed at the top of the rankings.

The utility and decision theory literature is rich and this decision theoretic analysis of the Tichy/Oddie method (and in general other similar methods) opens the door up to further experimentation and the application of decision theoretic resources to truthlikeness/information methods.

4 Another Method to Quantify Semantic Information

It is now time to propose a new method to measure information along truthlikeness lines. As will be seen, this approach differs to both the Tichy/Oddie and Niinniluoto approaches, lying somewhere in between. Once again, each state is assigned a value and ranked, where the value for a state w ($\text{val}(w)$) is determined using the following method:

- Let n stand for the number of propositional variables in the logical space
- Let t stand for the number of true atoms in a state w relative to the actual state
- $\text{val}(w) = \frac{t}{n \times 2^n}$

So in the case of our 3-atom logical space with w_1 being the actual state, each state is valued as follows:

- $\text{val}(w_1) = \frac{3}{24}$
- $\text{val}(w_2) = \text{val}(w_3) = \text{val}(w_5) = \frac{2}{24}$
- $\text{val}(w_4) = \text{val}(w_6) = \text{val}(w_7) = \frac{1}{24}$
- $\text{val}(w_8) = 0$

Now, given a statement A, its information yield is measured using the following algorithm:

1. Determine the set of states W in which A holds ($W = \{w \mid w \models A\}$)
2. Place the members of W into an array X_1 of length $|W|$ and order the members of X_1 from lowest to highest value

   This process is represented with a function arraystates(), so that $X_1 = \text{arraystates}(W)$.

3. Let X_2 stand for an empty array with 2^n spaces. Start by placing the first (lowest) element of X_1 in the first position of X_2. In general, place the n^{th} element of X_1 in the n^{th} position of X_2.

   Unless the statement being measured is a tautology, then length(X_1) $<$ length(X_2). So once the last element of X_1 has been reached, use this last element to fill in all the remaining places of X_2.

   This process is represented with a function lineup(), so that $X_2 = \text{lineup}(X_1)$.

4. Finally, sum up all the values of each element of X_2 to get the information measure

$$\text{info(A)} = \text{sum}(X_2) = \text{sum}(\text{lineup}(X_1)) = \text{sum}(\text{lineup}(\text{arraystates}(W))).$$

Following is an example of this method, which shall henceforth be referred to as the *value aggregate* method.

Example Let A be the statement h. Then:

- $W = \{w_1, w_2, w_3, w_4\}$
- $X_1 = [w_4, w_3, w_2, w_1]$
- $X_2 = [w_4, w_3, w_2, w_1, w_1, w_1, w_1, w_1]$
- $\text{info}(h) = \text{sum}(X_2) = val(w_4) + val(w_3) + val(w_2) + 5val(w_1) = 0.8333$

Remark With a ranking (highest to lowest) of all 255 possible statements (propositions, state collections), in the 3-atom space, the ordering of statements given by this method largely agrees with Niiniluoto's min-sum measure (with γ and λ as given in Section 3.5) for a significant portion of the first half of ordered statements. In fact it agrees for the first 107 positions, after which telling differences emerge. Also, it agrees more with the Tichy/Oddie approach towards the bottom of the rankings.

Table 5 contains some results using the value aggregate method. They seem reasonable and it is evident that their spread is closer to the Niiniluoto spread than the Tichy/Oddie. Also, unlike the Tichy/Oddie approach, where the differences between h, $h \wedge r$ and $h \wedge r \wedge w$ are constant, with the value aggregate approach the first of these differences is greater than the second.

#	Statement (A)	T/F	info(A)
1	$h \wedge r \wedge w$	T	1
2	$h \wedge r$	T	0.9583
3	$h \wedge (r \vee w)$	T	0.9167
4	$h \wedge (\neg r \vee w)$	T	0.875
5	h	T	0.8333
6	$(h \wedge r) \vee w$	T	0.7917
7	$(h \wedge \neg r) \vee w$	T	0.75
8	$h \vee r$	T	0.7083
9	$h \wedge r \wedge \neg w$	F	0.6667
10	$h \vee \neg r$	T	0.625
11	$h \vee r \vee w$	T	0.625
12	$h \wedge \neg r$	F	0.625
13	$h \vee \neg r \vee \neg w$	F	0.5833
14	$h \vee r \vee \neg w$	F	0.5833
15	$(h \vee \neg w) \wedge \neg r$	F	0.5417
16	$\neg h$	F	0.5
17	$h \vee \neg h$	T	0.5
18	$\neg h \vee \neg r \vee \neg w$	F	0.4583
19	$h \wedge \neg r \wedge \neg w$	F	0.3333
20	$\neg h \wedge \neg r$	F	0.2917
21	$\neg h \wedge \neg r \wedge \neg w$	F	0
22	$h \wedge \neg h$	F	N/A

Table 5: Information yield results using the value aggregate method

4.1 Adequacy Conditions

As mentioned earlier in Section 3.4, Niiniluoto states a number of adequacy conditions "which an explicate of the concept of truthlikeness should satisfy" [6, p. 232.]. An investigation into the applicability of these conditions to an account of semantic information quantification (whether all or only some apply) will not be pursued here; apart from perhaps an occasional comment, I do not intend to discuss matter. At the least most of them seem applicable.

Following is a list of these conditions plus a summary of how the Tichy/Oddie, Niiniluoto and value aggregate approaches fare against them. This summary confirms that the value aggregate approach is similar but not equivalent to either of the other two approaches.

The presentation of these conditions will largely conform to their original form, so bear in mind for our purposes that Tr() corresponds to info() and as Δ decreases/increases info() increases/decreases. Before listing the conditions, some terms need to be established:

- Niiniluoto uses the term *constituent* as a more general term for *state descriptions*.

- A is used to refer to statements in the logical space in general. $\mathbf{I}_A$ is the set of numbers used to index the set of states corresponding to the statement A. For example, in our weather framework, the statement h corresponds to the states $\{w_1, w_2, w_3, w_4\}$, so letting A stand for h we have $\mathbf{I}_A = \{1, 2, 3, 4\}$.
- S_i is used to refer to the constituent (state description) that corresponds to state w_i.
- S_* is reserved for the state description that corresponds to the actual state, so in our case $S_* = S_1$.
- $\mathbf{B}$ stands for the set of mutually exclusive and jointly exhaustive constituents. So in our case $\mathbf{B} = \{w_i \mid 1 \leq i \leq 8\}$. $\mathbf{I}$ is the set of numbers corresponding to $\mathbf{B}$.
- Δ_{ij} stands for $\Delta(w_i, w_j)$.
- An element S_j of $\mathbf{B}$ is called a Δ-*complement* of S_i, if $\Delta_{ij} = \max\Delta_{ik}$, $k \in \mathbf{I}$. Also, our weather framework example is Δ-*complemented*; if each $S_i \in \mathbf{B}$ has a unique Δ-*complement* S_j in $\mathbf{B}$, such that $\Delta_{ij} = 1$, the system $(\mathbf{B}, \Delta)$ is said to be Δ-*complemented.*

Here are the conditions:

(M1) ***(Range)*** $0 \leq \mathrm{Tr}(A, S_*) \leq 1$.

(M2) ***(Target)*** $\mathrm{Tr}(A, S_*) = 1$ iff $A = S_*$.

(M3) ***(Non-triviality)*** All true statements do not have the same degree of truthlikeness, all false statements do not have the same degree of truthlikeness.

(M4) ***(Truth and logical strength)*** Among true statements, truthlikeness covaries with logical strength.

  (a) If A and B are true statements and $A \vdash B$, then $\mathrm{Tr}(B, S_*) \leq \mathrm{Tr}(A, S_*)$.

  (b) If A and B are true statements and $A \vdash B$ and $B \nvdash A$, then $\mathrm{Tr}(B, S_*) < \mathrm{Tr}(A, S_*)$.

(M5) ***(Falsity and logical strength)*** Among false statements, truthlikeness does not covary with logical strength; there are false statements A and B such that $A \vdash B$ but $\mathrm{Tr}(A, S_*) < \mathrm{Tr}(B, S_*)$.

(M6) ***(Similarity)*** $\mathrm{Tr}(S_i, S_*) = \mathrm{Tr}(S_j, S_*)$ iff $\Delta_{*i} = \Delta_{*j}$ for all $S_i, S_j \in \mathbf{B}$.

(M7) (***Truth content***) If A is a false statement, then $\text{Tr}(S_* \vee A, S_*) > \text{Tr}(A, S_*)$.

(M8) (***Closeness to the truth***) Assume $j \notin \mathbf{I}_A$. Then $\text{Tr}(A \vee S_j, S_*) > \text{Tr}(A, S_*)$ iff $\Delta_{*j} < \Delta_{min}(A, S_*)$.

(M9) (***Distance from the truth***) Let $\Delta_{*1} < \Delta_{*i}$. Then $\text{Tr}(S_1 \vee S_j, S_*)$ decreases when Δ_{*i} increases.

(M10) (***Falsity may be better than truth***) Some false statements may be more truthlike than some true statements.

(M11) (***Thin better than fat***) If $\Delta_{*i} = \Delta_{*j} > 0$, $i \neq j$, then $\text{Tr}(S_i \vee S_j, S_*) < \text{Tr}(S_i, S_*)$.

(M12) (***Ovate better than obovate***) If $\Delta_{*1} < \Delta_{*i} < \Delta_{*2}$, then $\text{Tr}(S_1 \vee S_i \vee S_2, S_*)$ increases when Δ_{*i} decreases.

(M13) (**Δ-*complement***) $\text{Tr}(A, S_*)$ is minimal, if A consists of the $\Delta-$complements of S_*.

Table 6 gives a summary of the measures against the adequacy conditions, where $+(\gamma)$ means that the measure satisfies the given condition with some restriction on the value of γ and (λ) [13]:

condition	Niiniluoto (ms)	Tichy/Oddie (av)	Value Aggregate
M1	+	+	+
M2	+	+	+
M3	+	+	+
M4a	+	−	+
M4b	+	−	+
M5	$+(\gamma)$	+	+
M6	+	+	N/A
M7	+	+	+
M8	$+(\gamma)$	−	+
M9	+	+	+
M10	$+(\gamma)$	+	+
M11	+	−	−
M12	+	+	+
M13	$+(\gamma)$	+	+

Table 6: Measures against adequacy conditions

As can be seen, the value aggregate approach fails only M11. $\Delta_{*2} = \Delta_{*3} = \Delta_{*5}$, yet $\text{info}(w_2 \vee w_3 \vee w_5) = \text{info}(w_2 \vee w_3) = \text{info}(w_2)$. The value aggregate method, like the Tichy/Oddie method, does not differentiate sets of states such as these.

[13] See appendix for proofs relating to the value aggregate approach

4.2 Misinformation and Adjusting Utilities

Using the value aggregate method, a metric for misinformation measurements is once again simply obtained by inverting the utility values for each state; where u is original utility of a state, its new utility becomes $1 - u$. As with the Tichy/Oddie method, this is equivalent to $\text{misinfo}(A) = 1 - \text{info}(A)$. Unlike the Tichy/Oddie method though, where there is a perfect symmetry between truth and falsity, the accumulation of values with the aggregate approach results in a skew towards the information measure. As a consequence there are significant differences between the two in terms of misinformation value calculations. For example, whilst the Tichy/Oddie approach gives $\text{misinfo}(w_1 \vee w_8) = \text{misinfo}(w_4 \vee w_5) = 0.5$, this new approach gives $\text{misinfo}(w_1 \vee w_8) = 0.125 < \text{misinfo}(w_4 \vee w_5) = 0.375$.

Also, the value aggregate method can be extended to incorporate different utility functions for different purposes, in the same way as discussed in Section 3.6.1. As described in earlier Section 4, with the value aggregate approach the last element in the array X_1 is the highest valued and is used to fill in the remaining positions of X_2. In the case of w_1 in particular, this means that a collection of states consisting of w_1 relatively few other states will jump ahead quite quickly given the continued addition of the highest valued element.

The adoption of a non-linear utility function could be used to regulate this. For example, a simple logarithmic utility function such as $y = 20\log_2(x + 1)$ (x being the value of the standard utility) places w_1, w_8 below w_1, w_3, w_5, w_7, whereas with the original standard linear utility this order is reversed. Also to note, as long as the actual state is assigned the highest utility, the information measure will covary with logical strength for true statements, whatever utility function is used.

5 Combining CSI and truthlikeness approaches

I would like to now briefly discuss the possibilities of (1) using a CSI-style inverse probabilistic approach as a basis for a quantitative account of semantic information (2) combining the CSI approach and truthlikeness approaches.

With regards to (1), there seems to be inherent issues with the CSI inverse probabilistic approach. Thinking about the two approaches to semantic information quantification we have looked at (CSI and truthlikeness), Table 7 depicts a rough conceptual principle that can be extracted. The four possible ways of combining the two factors under consideration are represented using a standard 2x2 matrix, with each possible way being assigned a unique number between 1 and 4. 1 represents 'best' for information yield and 4 represents 'worst'.

	More Truth than Falsity	More Falsity than Truth
High Reduction of Possibilities	1	4
Low Reduction of Possibilities	2	3

Table 7: Possibility reduction and truth/falsity combinations

So a narrowing down of the range of possibilities is a good when done truthfully. On the other hand, if you are doing so falsely it is a bad thing. Borrowing an idea from decision theory, this type of reward/punishment system can be formally captured with a simple *scoring rule*. Take the following, where for a statement A:

- if A is true then give it an information measure of $\mathrm{cont}(A)$
- if A is false then give it an information measure of $-\mathrm{cont}(A)$

Whilst this approach would give acceptable results for true statements, when it comes to false statements it is too coarse and does not make the required distinctions between the different classes of false statements. A state description in which every atom is false is rightly assigned the lowest measure. Furthermore, other false statements which are true in the state corresponding to that false state description will be assigned acceptable measures.

But this approach also has it that any false state description will be accorded the lowest measure. So the state description $h \wedge r \wedge \neg w$ would be assigned the same measure as $\neg h \wedge \neg r \wedge \neg w$, which is clearly inappropriate. Furthermore, equal magnitude supersets of the states correlating to these state descriptions (e.g. w_2, w_4 and w_4, w_8) would also be assigned equal measures.

Given the aim to factor in considerations of truth value, the problem with any account of information quantification based on the CSI inverse probabilistic approach is related to the failure of Popper's content approach to truthlikeness. As Niiniluoto nicely puts it:

> among false propositions, increase or decrease of logical strength is neither a sufficient

> nor necessary condition for increase of truthlikeness. Therefore, any attempt to define truthlikeness merely in terms of truth value and logical deduction fails. More generally, the same holds for definitions in terms of logical probability and information content.[7, p. 296.]

With regards to (2), one option is to adopt a hybrid system. As was seen, the CSI approach works well when its application is confined solely to true statements. So perhaps a system which applies the CSI approach to true statements and a truthlikeness approach to false statements could be used. Furthermore, apart from such hybrid systems, there is the possibility of combining calculations from both approaches into one metric. For example, an incorporation of the CSI approach into the value aggregate approach could help distinguish between, say w_2 and w_2, w_3, which are given the same measure using the value aggregate approach.

6 Formula-Based Approaches

In *Updating Logic Databases*, classes of update semantics for logical databases are divided into statement-based and model-based approaches. Basically, with model-based approaches the semantics of an update for a database are based on the models of that database and with statement-based approaches statements in the database are operated on instead [12, p. 12.].

Applying this division to our investigation, it can be seen that the approaches discussed thus far would fall under a model-based approach to semantic information quantification. Beyond suggesting the possibility of a statement-based approach, I will also outline a method that can be seen to fall on this side of the divide. This outline is more so for the sake of illustration rather than anything else, though it could be useful.

To begin with, statements are first converted to conjunctive normal form. Although this means that logically equivalent statements are treated the same, they are still being dealt with directly and there is no analysis whatsoever of the models that they correspond to.

A logical statement is in conjunctive normal form (CNF) when it consists of a conjunction of disjunctive clauses, with each disjunct in each conjunction being a literal (either an atom or a negated atom). $A \wedge B \wedge \neg C$ and $(A \vee \neg B) \wedge (A \vee C)$ are two examples of statements in CNF.

Also, the normalised statements are *fully* converted to CNF, meaning that all redundancies are eliminated. Here is the procedure:

- use equivalences to remove $\leftrightarrow$ and $\rightarrow$
- use De Morgan's laws to push negation signs immediately before atomic statements
- eliminate double negations
- use the distributive law $A \vee (B \wedge C) \equiv (A \vee B) \wedge (A \vee C)$ to effect the conversion to CNF
- if a conjunct contains both A and $\neg A$, remove it
- use absorption law $A \wedge (A \vee B) \equiv A$ so that for any literal that occurs as a sole disjunct (i.e occurs in a disjunctive clause of which it is the only member), remove any others disjunctive clauses that contain that literal.
- (use rule of unsatisfiability $A \vee \perp \equiv A$ so that for any literal A that occurs as a sole disjunct (i.e occurs in a disjunctive clause of which it is the only member), remove its negation from any disjunction it occurs in

Given a normalised statement, how can its information yield be measured? Let us start with the simplest type of statement in CNF, one consisting of a conjunction of literals. Each literal has a value of magnitude 1. At this stage there are two ways that one can go. The first method (Method 1) involves just positively incrementing the total information yield of the statement as each true literal is encountered. The second method (Method 2) involves also decrementing the total information yield of the statement as each false literal is encountered.

Take the statement $h \wedge r \wedge \neg w$. It contains 2 true literals and 1 false literal. Going by the first method, the total information yield of this statement is 2. Going by the second method, where the misinformation instances of a statement count against its total information yield, the total information yield of this statement is 1.

This is all very simple and gives the right results. $h \wedge r \wedge w$ has a maximal information yield of 3. $h \wedge r \wedge \neg w$ and $h \wedge r$ have the same information yield of 2 going by Method 1, but going by method 2 the former is 'punished' for its assertion of misinformation and goes down to 1.

Method 1

#	Statement (A)	T/F	info(A)
1	$h \wedge r \wedge w$	T	3
2	$h \wedge r \wedge \neg w$	F	2
3	$h \wedge r$	T	2
4	$h \wedge (r \vee w)$	T	1.5
5	$h \wedge (\neg r \vee w)$	T	1.25
6	$h \wedge \neg r \wedge \neg w$	F	1
7	$h \wedge \neg r$	F	1
8	$h \wedge \neg h$	F	1
9	h	T	1
10	$(h \wedge r) \vee w$	T	1
11	$(h \wedge \neg r) \vee w$	T	0.75
12	$h \vee r$	T	0.5
13	$h \vee r \vee w$	T	0.33
14	$(h \vee \neg w) \wedge \neg r$	F	0.25
15	$h \vee \neg r$	T	0.25
16	$h \vee \neg h$	T	0.25
17	$h \vee r \vee \neg w$	T	0.22
18	$h \vee \neg r \vee \neg w$	T	0.11
19	$\neg h \wedge \neg r \wedge \neg w$	F	0
20	$\neg h$	F	0
21	$\neg h \wedge \neg r$	F	0
22	$\neg h \vee \neg r \vee \neg w$	F	0

Method 2

#	Statement (A)	T/F	info(A)
1	$h \wedge r \wedge w$	T	3
2	$h \wedge r$	T	2
3	$h \wedge (r \vee w)$	T	1.5
4	$h \wedge r \wedge \neg w$	F	1
5	$h \wedge (\neg r \vee w)$	T	1
6	h	T	1
7	$(h \wedge r) \vee w$	T	1
8	$(h \wedge \neg r) \vee w$	T	0.5
9	$h \vee r$	T	0.5
10	$h \vee r \vee w$	T	0.33
11	$h \vee \neg r$	T	0.25
12	$h \vee \neg h$	T	0
13	$h \wedge \neg r$	F	0
14	$h \wedge \neg h$	F	0
15	$h \vee r \vee \neg w$	T	0.11
16	$h \vee \neg r \vee \neg w$	T	-0.11
17	$\neg h \vee \neg r \vee \neg w$	F	-0.33
18	$h \wedge \neg r \wedge \neg w$	F	-1
19	$(h \vee \neg w) \wedge \neg r$	F	-1
20	$\neg h$	F	-1
21	$\neg h \wedge \neg r$	F	-2
22	$\neg h \wedge \neg r \wedge \neg w$	F	-3

Table 8: Information yield using formula-based approach

The introduction of proper disjunctions makes things more complicated. How can the calculation method reflect the fact that $h \vee r$ yields less information than other statements such as h and $h \wedge r$. The more disjuncts a disjunction contains, the less information it yields; information yield is inversely proportional to the number of disjuncts. Hence, where n is the number of disjuncts in a disjunct, a simple yet potentially suitable measure of the maximum information yield of the disjunction is $\frac{1}{n}$. With this way, each literal in the disjunction has a value of magnitude $\frac{1}{n^2}$.

To illustrate all of this, take the statement $h \vee r \vee w$. The total information yield of this statement is $3 \times \frac{1}{9} = \frac{1}{3}$. Given a disjunction with false literals, the results depend of whether Method 1 or Method 2 is adopted. For the statement $h \vee r \vee \neg w$, Method 1 gives a result of $\frac{1}{3} + \frac{1}{3} = \frac{2}{3}$ and Method 2 gives a result of $\frac{1}{3} + \frac{1}{3} - \frac{1}{3} = \frac{1}{3}$.

Table 8 lists results using Method 1 and results using Method 2.

7 Estimated Information

In closing I shall briefly introduce the idea of estimated information. Within the literature a distinction is made between the *semantic* and *epistemic* problems of truthlikeness [13, p. 121.]:

- *The semantical problem*: "What *do we mean* if we claim that the theory X is closer to the truth than the theory Y?"
- *The epistemic problem*: "On what evidence *are we to believe* that the theory X is closer to the truth than the theory Y?"

The focus of this paper has been the semantical problem of information quantification; what do we mean when we say that statement A yields more information than statement B? The epistemic problem relates to estimating information yield; given some partial evidence E, are we to estimate that statement A yields more information or less information than statement B?

Of course, an account of the semantical problem is primary, with any account of the epistemic problem being secondary, based on a semantical foundation. Nonetheless a method to estimate information yield is of the highest importance.

In practice judgements of information yield are going to be made with limited evidence, without knowledge of the complete truth (the one true state description). If an agent already knows the complete truth in a domain of inquiry then there would be no new information for them to acquire, no statement in the domain of inquiry could be informative for them. In such scenarios, the actual information yield of any statement is already known, or can be calculated.

In general though, calculations of information yield are of interest to agent's who do not know the complete truth and who are seeking information. When such agent's must choose amongst a set of different statements, their aim is to choose the statement that they estimate will yield the most information relative to the actual state (which the agent has limited epistemic access to). In making this estimation and choice the agent will often already posses some evidence and this evidence will rule out certain possible states from consideration in the calculations.

The standard formula for estimated utility in decision theory can be used to calculate the expected information yield of a statement A given prior evidence E:

$$\mathrm{info}_{est}(A|E) = \sum_{i=1}^{n} \mathrm{info}(A, S_i) \times \Pr(S_i|E)$$

n stands for the number of possible states in the logical space and S_i stands for the state description corresponding to state i.

Given a certain piece of information as evidence, the difference between actual and estimated information yield calculations for a statement can be marked. For example, with info() as the Tichy/Oddie method:

- $\mathrm{info}(h \wedge r \wedge w) = 1$
- $\mathrm{info}_{est}(h \wedge r \wedge w | h \vee \neg r \vee \neg w) = 0.48$
- $\mathrm{info}(\neg h \wedge \neg r \wedge \neg w) = 0$
- $\mathrm{info}_{est}(\neg h \wedge \neg r \wedge \neg w | h \vee \neg r \vee \neg w) = 0.52$

A Appendix

Adequacy conditions for the value aggregate method.

- Let W_A stand for the set of states satisfying statement A.
- w_T denotes the actual state.
- $S_* = T$.
- Consult Section 4 for further terminology.

Theorem A.1. $0 \leq \mathrm{Tr}(A, T) \leq 1$.

Proof. $X = \mathrm{lineup}(\mathrm{arraystate}(W_A))$. The lowest possible value for $\mathrm{sum}(X)$ is 0, when each item of X is the lowest valued state ($2^n \times 0 = 0$). The highest possible value for $\mathrm{sum}(X)$ is 1, when each item of X is the highest valued state ($2^n \times \frac{n}{n \times 2^n} = 1$). □

Theorem A.2. $\mathrm{Tr}(A, T) = 1$ iff $A = T$

Proof. $X = \mathrm{lineup}(\mathrm{arraystate}(W_A))$. $\mathrm{sum}(X) = 1$ iff each item of X is the highest valued state (i.e. w_T). Each item of X is the highest valued state iff the statement being measured is a state description of the actual state. □

Theorem A.3. All true statements do not have the same degree of truthlikeness, all false statements do not have the same degree of truthlikeness.

Proof. Evident with the results from Table 5. □

Theorem A.4. Among true statements, truthlikeness covaries with logical strength:

(a) If A and B are true statements and $A \vdash B$, then $\mathrm{Tr}(B, T) \leq \mathrm{Tr}(A, T)$.

(b) If A and B are true statements and $A \vdash B$ and $B \nvdash A$, then $\mathrm{Tr}(B, T) < \mathrm{Tr}(A, T)$.

Proof. We will show that (b) holds, since this entails (a).

- $X_1^A = \text{arraystates}(W_A)$
- $X_1^B = \text{arraystates}(W_B)$
- $X_2^A = \text{lineup}(X_1^A)$
- $X_2^B = \text{lineup}(X_1^B)$

Now

- where n is the number of propositional variables, the values of the states will be drawn from the following: $\frac{0}{(n \times 2^n)}, \frac{1}{(n \times 2^n)}, \frac{2}{(n \times 2^n)} \cdots \frac{n}{(n \times 2^n)}$.
- Since A and B are true, $w_T \in W_A$ and $w_T \in W_B$.
- since $A \vdash B$ and $B \nvdash A$, $W_A \subset W_B$

So X_2^A is going to contain at least 1 more instance of w_T than X_2^B. Say that X_2^B does just have 1 less instance of w_T and that this is replaced by an instance of the second highest valued state (which will have a value of $\frac{n-1}{((n-1) \times 2^n)}$). This is the upper limit, so it will suffice to show that in this case $\text{sum}(X_2^B) < \text{sum}(X_2^A)$.

The term $X[i]$ denotes position i of array X. Let m be the position at which $X_2^A[m] \neq X_2^B[m]$. So $1 \leq m \leq 2^n - 1$ and for each $i \in \{x \mid x < m \wedge x \in [1, m-1]\}$, $X_2^A[i] = X_2^B[i]$, so the value of sum() is equal for both arrays up to point m.

After and including point m, whilst the remaining elements of X_2^A sum up to $(2^n - m) \times \frac{n}{(n \times 2^n)}$, the remaining elements of X_2^B sum up to $\frac{n-1}{(n \times 2^n)} + ((2^n - m) - 1) \times \frac{n}{(n \times 2^n)}$.

So the final thing to show is that $(2^n - m) \times \frac{n}{(n \times 2^n)} > \frac{n-1}{(n \times 2^n)} + ((2^n - m) - 1) \times \frac{n}{(n \times 2^n)}$

$$
\begin{aligned}
\frac{n(2^n - m)}{\cancel{(n \times 2^n)}} &> \frac{(n-1) + n(2^n - m - 1)}{\cancel{(n \times 2^n)}} \\
n2^n - nm &> n2^n - nm - 1
\end{aligned}
$$

$\square$

Theorem A.5. Among false statements, truthlikeness does not covary with logical strength; there are false statements A and B such that $A \vdash B$ but $\mathrm{Tr}(A, T) < \mathrm{Tr}(B, T)$.

Proof. Evident with the results from table 5. $\square$

Theorem A.6. If A is a false statement, then $\mathrm{Tr}(T \vee A, T) > Tr(A, T)$

Proof. Since A is false, it follows that $w_T \notin W_A$. The set of states corresponding to $T \vee A$ is $W_A \cup \{w_T\}$. Let

- $X_2 = \mathrm{lineup}(\mathrm{arraystates}((W_A)))$
- $X_{2'} = \mathrm{lineup}(\mathrm{arraystates}((W_A \cup \{w_T\})))$

So we need to show that $\mathrm{sum}(X_{2'}) > \mathrm{sum}(X_2)$.

Say the highest valued element of W_A is w_a and that w_a takes up the last n positions in X_2. The addition of w_T results in it replacing w_a for the last $n-1$ positions. Since $\mathrm{val}(w_T) > \mathrm{val}(w_a)$, $\mathrm{sum}(X_{2'}) > \mathrm{sum}(X_2)$. $\square$

Theorem A.7. Assume $j \notin \mathbf{I}_A$. Then $\mathrm{Tr}(A \vee S_j, T) > \mathrm{Tr}(A, T)$ iff $\Delta_{*j} < \Delta_{min}(A, T)$

Proof. First we show that

$$\Delta_{*j} < \Delta_{min}(A, T) \to \mathrm{Tr}(A \vee S_j, T) > \mathrm{Tr}(A, T)$$

If the antecedent here holds, then $(\forall w)(w \in W_A \supset \mathrm{val}(w_j) > \mathrm{val}(w))$, so adding w_j to the array will result in an ordering of states such that w_j is the last, rightmost. This means that the last value being added is higher than the one it replaced, resulting in a higher overall number.

Second we show that

$$\mathrm{Tr}(A \vee S_j, T) > \mathrm{Tr}(A, T) \rightarrow \Delta_{*j} < \Delta_{min}(A, T)$$

via the contraposition

$$\Delta_{*j} > \Delta_{min}(A, T) \rightarrow \mathrm{Tr}(A \vee S_j, T) < \mathrm{Tr}(A, T)$$

Since w_j has a value less than that of the highest valued element $w_a \in W_A$, adding it to the array lineup(W_A) means that one instance of w_a is taken off from the sum() calculations and replaced by a lower value, resulting overall in a lower value.

□

Theorem A.8. Let $\Delta_{*1} < \Delta_{*i}$. Then $\mathrm{Tr}(S_1 \vee S_j, T)$ decreases when Δ_{*i} increases

Proof. This holds straightforwardly. There are two states, w_1 which corresponds to S_1 and w_j which corresponds to S_j, with the value of w_1 being greater than the value of w_j. Let $X_2 =$ lineup(arraystates($\{w_1, w_j\}$)), so info($S_1 \vee S_j$) = sum(X_2). Let $X_{2'} =$ lineup(arraystates($\{w_{1'}, w_j\}$)), where $w_{1'}$ replaces w_1 when Δ_{*1} increases. Since w_1 is replaced by a lower $w_{1'}$, then sum($X_{2'}$) < sum(X_2). □

Theorem A.9. Some false statements may be more truthlike than some true statements.

Proof. Evident with the results from table 5. □

Theorem A.10. If $\Delta_{*1} < \Delta_{*i} < \Delta_{*2}$, then $Tr(S_1 \vee S_i \vee S_2, T)$ increases when Δ_{*i} decreases.

Proof. This is straightforward; any increase for $Tr(h_i)$ means an increase in the value of w_i, which means an overall total increase in sum(). □

Theorem A.11. $\mathrm{Tr}(A, T)$ is minimal, if A consists of the $\Delta-$complements of T.

Proof. If A is the Δ-complement of S_* then, it is the state such that each atom is false. Therefore, it adds up to 0. $\square$

References

[1] Yehoshua Bar-Hillel and Rudolf Carnap, 'Semantic Information'. The British Journal for the Philosophy of Science, Vol. 4. No. 14, 1953, 147157.

[2] Floridi, Luciano, 'Outline of a Theory of Strongly Semantic Information'. Minds and Machines, Vol. 14, 2004, 197221.

[3] Floridi, Luciano, 'Is Semantic Information Meaningful Data?'. Philosophy and Phenomenological Research, Vol. LXX, 2005, 351370.

[4] Floridi, Luciano, 'In Defence of the Veridical Nature of Information'. European Journal of Analytic Philosophy, Vol. 3. No. 1, 2007, 3141.

[5] Floridi, Luciano, 'Semantic Conceptions of Information', The Stanford Encyclopedia of Philosophy (Summer 2009 Edition), Edward N. Zalta (ed.), URL = ¡http://plato.stanford.edu/archives/sum2009/entries/information-semantic/¿.

[6] Niiniluoto, I. *Truthlikeness*, Dordrecht: Reidel, 1987.

[7] Niiniluoto, I. 'Review of Likeness to the Truth'. The Journal of Symbolic Logic, Vol. 54, No. 1 (Mar., 1989), pp. 296-297

[8] Oddie. G. *Likeness to Truth*, (Western Ontario Series in Philosophy of Science), Dordrecht: Reidel.

[9] Oddie, Graham, "Truthlikeness", The Stanford Encyclopedia of Philosophy (Fall 2008 Edition), Edward N. Zalta (ed.), URL = ¡http://plato.stanford.edu/archives/fall2008/entries/truthlikeness/¿.

[10] Priest, Graham. *An Introduction to Non-Classical Logic* (Volume I), Cambridge, Cambridge University Press, 2001.

[11] From SEP entry on semantic information

[12] Windslett, Marianne. *Updating Logical Databases*

[13] Zwart, Sjoerd D. *Refined Verisimilitude*, Kluwer, Dordrecht 2001