Outline of a Theory of Strongly Semantic Information

L. Floridi

Sub-Faculty of Philosophy and Computing Laboratory

University of Oxford

Wolfson College

Oxford, OX2 6UD, UK

luciano.floridi@philosophy.oxford.ac.uk

Outline of a Theory of Strongly Semantic Information

Abstract

This paper outlines a quantitative theory of strongly semantic information (TSSI) based on truth-values rather than probability distributions. The main hypothesis supported in the paper is that (i) the classic quantitative theory of weakly semantic information (TWSI) is based on probability distributions because (ii) it assumes that truth-values supervene on information, yet (iii) this principle is too weak and generates a well-known semantic paradox, whereas (iv) TSSI, according to which information encapsulates truth, can avoid the paradox and is more in line with the standard conception of what counts as information. After a brief introduction, section two outlines the semantic paradox entailed by TWSI, analysing it in terms of an initial conflict between two requisites of a quantitative theory of semantic information. In section three, three criteria of information equivalence are used to provide a taxonomy of quantitative approaches to semantic information and introduce TSSI. In section four, some further desiderata that should be fulfilled by a quantitative TSSI are explained. From section five to section seven, TSSI is developed on the basis of a calculus of truth-values and semantic discrepancy with respect to a given situation. In section eight, it is shown how TSSI succeeds in solving the paradox. Section nine summarises the main results of the paper and indicates some future developments.

1. Introduction

"A triangle has four sides": according to the classic theory of semantic information, there is more *semantic content*[1] in this contradiction than in the contingently true statement "the earth has only one moon". Bar-Hillel and Carnap [1953][2] were among the first to make explicit this prima facie counterintuitive inequality:

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 emphasized 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. (p. 229).

With a little hyperbole, we may conveniently refer to it as the Bar-Hillel-Carnap semantic Paradox (BCP).

Since its formulation, BCP has been recognised as an unfortunate, yet perfectly correct and inevitable consequence of any *theory of weakly semantic information* (TWSI; more on TWSI in section 2). As a consequence, the problem has often been either ignored[3] or tolerated[4] as the price of an otherwise valuable approach. Sometimes, however, attempts have been made to circumscribe its counterintuitive consequences. This has happen especially in Information Systems Theory (Winder [1997])—where consistency is an essential constraint that must remain satisfied for a database to preserve data integrity—and in Decision Theory, where inconsistent information is obviously of no use to a decision maker. In these cases, whenever there are no possible models that satisfy a statement or a theory, instead of assigning to it the maximum quantity of information, three strategies have been suggested:

1) assigning to all inconsistent cases the same, infinite information value (Lozinskii [1994]). This is in line with an economic approach, which defines x as impossible if and only if that x has an infinite price;

2) eliminating all inconsistent cases a priori from consideration, as impossible outcomes in decision-making (Jeffrey [1990]). This is in line with the syntactic approach developed by the *statistical theory of signals transmission*[5] (STST, more on this in a moment);

3) assigning to all inconsistent cases the same zero information value (Mingers [1997], Aisbett and Gibbon [1999]).

The latter approach is close to the one developed in this paper. Informally, the general hypothesis is that BCP indicates that something has gone essentially amiss with TWSI. TWSI is based on a semantic principle that is too weak, namely that truth-values supervene on information (see the quotation above). A semantically stronger approach, according to which information encapsulates truth, can avoid the paradox and is more in line with the ordinary conception of what counts as information.[6]

The actual viability of an alternative to TWSI is, of course, another matter. In this direction, however, STST already provides some initial reassurance. STST identifies the quantity of information associated with, or generated by, the occurrence of a signal (an event or the realisation of a state of affairs) with the elimination of possibilities (reduction in uncertainty) represented by that signal (event or state of affairs).[7] Now in STST no counterintuitive inequality comparable to BCP occurs, and the line of argument in this paper will be that, as in the case of STST, a *theory of strongly semantic information* (TSSI), based on alethic and discrepancy values rather than probabilities, can also successfully avoid BCP, although for quite different reasons.

Before developing TSSI, a note on the terminology, the concepts and the assumptions used in the paper is in order. In what follows, the term "infon" and the symbol σ (Devlin [1991]) will be used to refer to discrete items of information, irrespective of their semiotic code and physical implementation. This is in line with common practice in AI, Computer Science and ICT (information and communication technology), where the expression

"information resources" is used to refer to information in different formats, e.g. hard copy or digital texts, graphics, maps, tabular data etc. (Heck and Murtagh [1993]). Some basic concepts belonging to situation logic will also be adopted, still following work done by Devlin [1991], who articulates and defends a theory of semantic information, understood extensionally, in terms of situation logic principles first developed by Barwise and Perry [1983]. Finally, the following three principles of communication theory will be assumed without further justification:

a) every σ-source that generates, sends or transmits σ is treated as a *bona fide* source of information;

b) when it appears that σ can have a higher or lower degree of informativeness, it is attributed the highest of such degrees;

c) the channel of communication of σ is treated as ideally noiseless.

The primary aim of these principles is to shift the burden of proving that σ is not maximally informative from the sender to the receiver. They make it possible to set aside both error analysis issues and problems of a sceptical nature that would be out of place in this context.

2. The Bar-Hillel-Carnap Paradox

According to TWSI, the semantic content (CONT) of an infon σ can be identified *negatively*[8] with the set of all the descriptions of the possible states of the universe that are excluded by σ:

1   CONT(σ) =def. the set of all state-descriptions inconsistent with σ

Suppose E is a probabilistic experiment[9] in which infons are messages formulated by a source, using the standard language of set theory and classic, first-order logic with identity, to describe a domain of three individual constants D = {a, b, c} and 2 predicates either affirmed or negated {G, H}, in such a way that each individual (e.g. each figure of the plane) is either

5

G or ¬G (e.g. four-sided or not), and either H or ¬H (e.g. right-angled or not). E allows the formulation of 12 different atomic messages, according to the scheme $\Phi(x)$. Atomic messages are incomplete state-descriptions of E. Complex messages can have any length, but a complete state-description is a message σ consisting of a conjunction of 6 different and consistent atomic messages. Each conjunctive message σ denotes one of the n possible states of E's universe. Since in E the number s of types of predicates is 2 and the length l of a message σ is 6, the number of distinguishable messages is $s^l = 64$.

E is a microworld whose total possibility is 1. Its fixed and finite sample space can be described either ontologically, as the set of all possible states $W = \{w_1, w_2, \ldots, w_{64}\}$, or semantically, as the set of all jointly exhaustive and mutually exclusive messages $\Sigma = \{\sigma_1, \sigma_2, \ldots, \sigma_{64}\}$ (see Table 1).[10] Three direct consequences of [1] are that:

i) the inclusion of any other atomic message in any $\sigma_i \in \Sigma$ would necessarily result in σ_i becoming inconsistent. Thus, Bar-Hillel and Carnap [1953] describe Σ as the set of all, most strongly synthetic messages in E. Each σ_i is inconsistent with any other message in Σ, and any other conjunctive message that logically implies, or is stronger than, any σ_i is self-contradictory;

ii) the set of all messages inconsistent with a tautology is empty, therefore any tautology ($\top$) has minimum semantic content,[11] that is

2 $$\text{CONT}(\top) = \text{MIN}$$

iii) since any message is inconsistent with a contradiction ($\bot$), any contradiction has maximum semantic content,[12] that is

BCP $$\text{CONT}(\bot) = \text{MAX}$$

In BCP, two fundamental requisites of a quantitative theory of semantic information appear to be in conflict. In section 7, we shall see that their relation is actually somewhat more

complex, but at the moment it is sufficient to analyse it in its simplest form. The first requisite (R.1) states that $\mathrm{CONT}(\sigma)$ should be inversely related to $p(\sigma)$, the logical probability of σ. The "mathematically simplest relation" (Bar-Hillel and Carnap [1953], 302) fulfilling this requirement is the complement of 1:

R.1 $$\mathrm{CONT}(\sigma) = 1 - p(\sigma)$$

TWSI implements R.1 but, as an alternative, R.1 could also be expressed as an inverse proportion:

R.1* $$\mathrm{CONT}(\sigma) \propto k/p(\sigma)$$

here k is a constant of proportionality independent of the two values. This is the standard approach implemented by STST, where the quantity of information conveyed by σ is equivalent to the reciprocal of $p(\sigma)$, expressed in bits:

3 $$\mathrm{I}(\sigma) = \log_2 1/p(\sigma) = - \log_2 (\sigma)$$

The problem with implementing R.1* in TWSI is that $\mathrm{CONT}(\sigma) \times p(\sigma) = k$, and the equation would require a more complex treatment of $\mathrm{CONT}(\top)$ and $\mathrm{CONT}(\bot)$ as two limits of the continuous function $f\,(p(\sigma))$:

4 $$\mathrm{CONT}\,(\top) = \lim_{p(\sigma) \to 1} f(p(\sigma)) = 0$$
$$\mathrm{CONT}(\bot) = \lim_{p(\sigma) \to 0} f(p(\sigma)) = 1$$

This complication does not occur in STST because the latter presupposes a frequency interpretation of probability and partitions the total amount of probability ($= 1.0$) a priori, among all the actually possible alternatives, thus excluding by definition the occurrence of any option that is necessarily false. In what follows, it is argued that the set-theoretic approach expressed in R.1 can be successfully adopted by TSSI as well.

The second requisite states that, *ceteris paribus*, $\iota(\sigma)$, i.e. the degree of informativeness of σ, should be directly proportional to the semantic content of σ:

7

R.2* $\iota(\sigma) \propto m \, \text{CONT}(\sigma) + b$, with $m \geq 1$ and $b \geq 0$

Since for our purposes the significant point in R.2* is only the relation of direct proportionality between $\iota(\sigma)$ and $\text{CONT}(\sigma)$, for the sake of elegance and simplicity we can adopt a weaker version of R.2*, by treating m and b as redundancy factors and hence reducing m to 1 and b to 0. Assuming that σ can be fully normalised in this way,[13] R.2* simplifies to:

R.2 $\iota(\sigma) \propto \text{CONT}(\sigma)$

Anything that will be inferred from R.2 below can be inferred, *a fortiori*, from R.2*.

R.1 and R.2 generate no conflict about the interpretation of $\models \sigma$. However, regarding $\sigma \models$, the conclusion is that $p(\sigma) = 0$ and $\iota(\sigma) = 0$, and it becomes unclear whether the value of $\text{CONT}(\sigma)$ should be MAX, following R.1, or MIN, following R.2. This tension is sufficiently problematic to invite the elaboration of a different approach.

3. Three criteria of information equivalence

TWSI, the classic quantitative theory of semantic information, concentrates on the (degree of) systemic consistency and then the a priori, logical probability of (sets of) infons. There is no reference to the actual alethic values of the infons in question, which are supposed to qualify as instances of information independently of whether they are true or false. This is why the theory can be described as only weakly semantic. Is it possible to avoid BCP by assuming a stronger semantic principle, according to which if σ qualifies as information it must encapsulate truth? The question presupposes a clear view of what alternatives to TWSI are available, hence a taxonomy of quantitative theories. The latter can be provided on the basis of three criteria of semantic equivalence.

Any quantitative theory, including any theory dealing with the concept of semantic information and its various measures (Smokler [1966]), requires at least a criterion of

comparative quantitative equivalence. In general, this "scale" criterion makes it possible to establish whether two measured objects, e.g. two coins x and y, have the same weight z, even if the latter cannot be qualified any more precisely. Now, two infons $\sigma_n \neq \sigma_m$ can be said to be co-informative— i.e. to possess an equivalent quantity of semantic information, Ci (σ_n, σ_m)—according to three criteria of comparative quantitative equivalence. They can be listed here in order of decreasing strength:

C.1   Ci $(\sigma_n, \sigma_m) \leftrightarrow \sigma_n$ and σ_m have equivalent meaning

It is reasonable to suppose, for example, that declarative sentences expressing the same proposition—like "Peter drives the car" and "The car is driven by Peter"—possess an equivalent amount of semantic information. If σ_n and σ_m are *qualitatively* co-informative, a *pragmatic theory* (Bar-Hillel and Carnap [1953]) could provide the relevant analysis, by addressing the question of how much information a certain infon carries *for a subject* S in a given doxastic state and within a specific informational environment. The pragmatic theory of "interested information" is crucial in Decision Theory, where a standard quantitative axiom states that, in an ideal context and *ceteris paribus*, the more informative σ is to S, the more S ought to be rationally willing to pay to find out whether σ is true (Sneed [1967]). It remains to be seen whether a satisfactory quantitative theory can effectively be developed in full.[14]

C.2   Ci $(\sigma_n, \sigma_m) \leftrightarrow \sigma_n$ and σ_m are truth-functionally equivalent

If σ_n and σ_m are *alethically* co-informative, Boolean algebra provides the relevant analysis. Of course C.2 can be extended to apply to classic n-order, multi-valued or fuzzy logic systems, etc.

C.3   Ci $(\sigma_n, \sigma_m) \leftrightarrow \sigma_n$ and σ_m are equiprobable

If σ_n and σ_m are *quantitatively* co-informative, the classic theory of weakly semantic information provides the relevant analysis.

Interpreting C.1-C.3 as conditionals rather than biconditionals provides the ground for a taxonomy of theories (see Table 2). More finely-grained criteria for co-informativeness could now be developed, ranging from full synonymity to full equiprobability. Intuitively, the closer a criterion is to synonymity, the more the semantic theory is forced to rely on the interpretation of σ by an intelligent receiver S capable of understanding its contextual sense (this is known as "fully interested information"), the less easily the information in question can undergo a quantifiable treatment, and the more a hermeneutic approach becomes inevitable. At the other end of the spectrum, a purely quantitative approach to the theory of semantic information will tend to abstract from the users of σ ("fully disinterested information") and to deal only with the analysis of $p(\sigma)$, irrespective of its *actual* interpretation and contextual alethic value. The latter strategy is consistent with a family of realist positions, including mathematical Platonism and epistemological Cartesianism: infons are treated extensionally not intentionally, as semantic (at least in the sense of being interpretable), structured, abstract objects, comparable to fundamental informational particles. They subsist, are systemically related and may be physically implemented in the world of human experience. [15]

A cursory analysis of Table 2 already suffices to show that:

a) approach no. 8 is trivially uninteresting;

b) approach no. 3 is not implementable, since it is impossible for Ci (σ_n, σ_m) to be the case qualitatively without σ_n and σ_m also being alethically co-informative;

c) approaches nos. 1, 2 and 4 would initially rank as the most interesting alternatives, but a fully quantitative and extensionalist approach to the meaning of σ seems unachievable, witness the failure of many programs in "strong" Artificial Intelligence (Floridi [1999],

chapter 5). In this paper it will be assumed that, if possible, less demanding alternatives should be developed first;

d) approach no. 6 is the one followed in mathematical logic; and

e) approach no. 7 is the one followed by TWSI.

We have seen that the implementation of C.3 in no. 7 brings with it some obvious computational advantages, of which TWSI makes the most, but also a shortcoming, namely BCP. The avoidance of BCP invites the development of

f) approach no. 5, that is an analysis of the quantity of semantic information in σ including a reference to its truth-value. This is TSSI.

4. Three desiderata for TSSI

As anticipated above, the principal goal of TSSI is to provide a quantitative analysis of semantic information that can be as mathematically successful as TWSI, but more respectful of our common-sense intuitions. More specifically, this means trying to:

D.1   avoid any counterintuitive inequality comparable to BCP

D.2   treat the alethic value of σ not as a supervenient but as a necessary feature of
             information, relevant to the quantitative analysis (for a similar approach see
             Dretske [1981])

D.3   extend a quantitative analysis to the whole family of information-related
             concepts: information   vacuity   and   inaccuracy,   informativeness,
             misinformation (what is ordinarily called "false information"), disinformation.

Arguably, D.1 and D.2 are complementary. BCP-like problems can arise because the alethic values of the infons in question are treated as irrelevant to the quantitative analysis of information, so TSSI can attempt to fulfil both D.1 and D.2 by implementing C.2. Regarding D.3, this requires a more robust and substantial theory of semantic information than one

based only on probability distributions (Szaniawski [1984], Mingers [1997]). In sections 6-9 it will be shown that, by treating information as encapsulating truth, TSSI offers a quantitative approach that can be generalised to a whole family of related concepts, although in this paper the quantitative analyses of misinformation and disinformation will not be provided for reasons of space (see section 9).

5. Degrees of vacuity and inaccuracy

The first step in the construction of TSSI is to define the concept of "informative content" or intrinsic *informativeness* of σ extensionally and a priori in an ideal context,[16] as a function of the positive or negative degree of "semantic distance" or *deviation* of σ from a fixed point or origin, represented by the given situation w, to which σ is supposed to refer.[17] Let w be a situation in a context, where a "situation" is "determined by (what goes on in) a topologically simply-connected, structured region of space-time" (Devlin [1991], 69), and a "context" is the set of interrelated conditions in which a situation occurs, what can be described, informally, as the immediate environment of a situation or, topologically, its neighbourhood. Each message σ_i in E conforms to (has the property of providing or conveying true contents, e.g. facts or ideas about) its corresponding situation w_i in W. Following the terminology of situation logic, this means that w_i fully supports σ_i (in symbols, $\Vdash_{w_i} \sigma_i$). Obviously, each message σ_i is maximally informative relatively to its situation w_i, yet this applies indiscriminately and trivially to any message. More interesting is the fact that the amount of informativeness of each σ_i can be evaluated absolutely, as a function of

a)   the polarity of σ_i, i.e. the alethic value possessed by σ_i; and

b)   the degree of discrepancy[18] (want of agreement) between σ_i and a given state of the world w, calculated as the degree ϑ of semantic deviation[19] of σ_i from the uniquely determinate state w in which E is ($\Vdash_{w\theta} \sigma$).

The given situation w functions as a benchmark to test the Boolean truth-value and the degree of discrepancy of each σ. This captures an important aspect of the ordinary conception of the nature of information, at the roots of any epistemic utility function (Levi [1967]): two infons can both be false and yet significantly more or less distant from the event or state of affairs w about which they purport to be informative (e.g. "there are ten people in the library" and "there are fifty people in the library", when in fact there are nine people in the library). Likewise, two infons can both be true and yet deviate more or less significantly from w (e.g. "there is someone in the library" vs. "there are 9 or 10 people in the library"). This implies that a falsehood with a very low degree of discrepancy may be pragmatically preferable to a truth with a very high degree of discrepancy (Popper [1959]). This in turn provides a strong argument for the rejection of the deductive closure principle: since σ could be false, if S commits him/herself to accepting σ as epistemically preferable, S is *not* necessarily committing him/herself to accepting *all* the deductive consequences of σ as epistemically preferable.

In order to express both positive (for σ = true) and negative (for σ = false) degrees of discrepancy, let $f(\sigma)$ be a mapping function from members of Σ to members of the set of some numeric values in the range of real numbers $[-1.0, +1.0]$. The function associates a real value ϑ of discrepancy to each σ depending on its truth-value and deviation from w:

5                    $\vartheta = f(\sigma)$

The mapping generates a continuous set Γ of ordered pairs $<\sigma, \vartheta>$. Γ is a subset of the Cartesian product $\Sigma \times [-1.0, +1.0]$ that may be equipotent with respect to Σ.[20] In a way that only partly resembles what happens in fuzzy logic, the lower and upper bound are used to represent complete discrepancy, while a selection of values in between is used to represent what are, intuitively, degrees of approximation to w from the negative and the positive side, with 0 as the indication of no discrepancy, i.e. complete conformity (note that ϑ values are

mapped onto the x-axis in the graph in Table 5 since values on the y-axis are the result of a composite function).

Intuitively, ϑ indicates the distance of an infon σ from a selected situation w and can be read as the degree of support offered by w to σ. In a condition of total ignorance, the value of ϑ can be established a priori by using a non-monotonic or probabilistic calculus (this is what justifies the clause "it is estimated that" in the following formulae), yet ϑ values should not themselves be interpreted as probabilities. Although they can have forecasting significance and may be based on the statistical strength of evidence supporting the assertion that σ conforms to w, they can be negative and they do not need to sum to one. In real-life cases, it may be difficult to calculate ϑ values with a fully reliable degree of precision and approximations may often be inevitable. However, any feasible and satisfactory metric will have to satisfy the following five conditions:

M.1 $\Vert\text{-}_w\, \sigma \rightarrow f(\sigma) = 0$

if (it is estimated that) σ is *true* and conforms to w most *precisely* (i.e. in the least vacuous way) and *accurately*, it is assigned a zero degree of discrepancy, i.e. $\vartheta = 0$. This is the technical sense in which precise and accurate information can be seen as the threshold between vacuous truth and inaccurate falsehood.

M.2 $\Vert\text{-}_{\forall w}\, \sigma \rightarrow f(\sigma) = 1$

if (it is estimated that) σ is true and conforms to any situation, σ is a tautology and it is assigned a maximum degree of *positive* discrepancy. Every situation supports σ, so no other true σ could be more distant from w than a tautological σ. In this case, σ is described as having the highest degree of *semantic vacuity*, i.e. $\vartheta = 1$.[21]

M.3 $\Vert\text{-}_{\neg\exists w}\, \sigma \rightarrow f(\sigma) = -1$

if (it is estimated that) σ is false and conforms to no possible situation, σ is a contradiction and it is assigned a maximum degree of *negative* discrepancy. No possible situation supports σ, so no other false σ could be more distant from w than a contradictory σ. In this case, σ is described as having the highest degree of *semantic inaccuracy*,[22] i.e. $\vartheta = -1$.

M.4 $\qquad\qquad \Vdash_{\text{-w}-\theta} \sigma \to (0 > f(\sigma) > -1)$

if (it is estimated that) σ is contingently false, it is assigned a degree of discrepancy with a value less than 0 but greater than -1 (degrees of semantic inaccuracy).

M.5 $\qquad\qquad \Vdash_{\text{-w}+\theta} \sigma \to (0 < f(\sigma) < +1)$

if (it is estimated that) σ is contingently true but does not conform to w with the highest degree of precision, it is assigned a degree of discrepancy with a value greater than 0 but less than $+1$ (degrees of semantic vacuity).

According to M.1-M.5, σ_n in language L_n is semantically equivalent to σ_m in language L_m if and only if σ_n and σ_m have the same absolute alethic value and a comparable L-dependent degree of discrepancy with respect to w.

The implementation of M.1-M.3 causes no great difficulties but, when one deals with models that are not easily formalisable, the behaviour of the function $f(\sigma)$ in M.4-M.5 may become a matter of more conventional stipulations, which can grade ϑ values according to their comparative degree of vacuity or inaccuracy, relative to a specific application. There may be various ways of developing this comparative analysis. Since the present task is to provide a general outline of a strongly semantic theory, in this article the analysis can be limited to a paradigmatic model, which any real-life application will try to approximate. Let us consider M.4 first, since it is simpler than M.5.

According to M.4, the relative "amount of falsehood" in each contingently false σ indicates the *degree of inaccuracy* of σ with respect to w. This is then calculated as the ratio between the number of erroneous[23] atomic messages e in σ and its length:

Inac $\qquad\qquad\qquad -\vartheta(\sigma) = -e(\sigma)/l(\sigma)$

Inac allows us to partition $\Sigma = s^l$ into l disjoint classes of inaccurate σ {Inac$_1$, …, Inac$_l$}, and map each class to its corresponding degree of inaccuracy (see Table 3 for an application to the model E).

Consider now M.5. The *degree of vacuity* cannot refer to members of Σ immediately, but only to their logical transformations σ_j possessing a positive alethic value. Since each σ_j must be assumed to be true by definition, the analysis of e/l can no longer provide a clear measure of $\vartheta(\sigma)$. For example, both the set of complements or negations—$\Sigma^n = \{\neg\sigma_1, \neg\sigma_2,$ …, $\neg\sigma_{64}\}$—and the logically equivalent set of duals[24]—$\Sigma^\Delta = \{\sigma_1{}^\Delta, \sigma_2{}^\Delta, …, \sigma_{64}{}^\Delta\}$—of all strongly synthetic messages contain formulae with e = 0 that would count as having the absolutely lowest degree of vacuity, i.e. $\vartheta = 0$, despite the obvious fact that, since they contain disjunctions, they would conform to a number of situations, including w, greater than one. Similar difficulties arise in the development of any analysis based on the ratio between the number of correct atomic messages in σ and its length. Is there any other reliable criterion to quantify $\vartheta(\sigma)$?

Recall that a tautology has $\vartheta = +1$ because it is consistent with every possible situation. In the case of the most weakly synthetic messages in E, i.e. members of the two sets Σ^n and Σ^Δ, each $\neg\sigma_i$ and $\sigma_i{}^\Delta$ is inconsistent with only one state in W. This means that members of the two sets have the highest possible degree of positive discrepancy to w, short of being analytically true. The conclusion can be generalised to any set of not strongly synthetic messages Σ^x. We shall say that, when σ is contingently true but vacuous, its semantic

distance from w can be calculated as a function of the number of situations, including w itself, with which σ is consistent:

Vac $\qquad\qquad\qquad\qquad +\vartheta(\sigma) = n/s^{l}$

More precisely, the degree of semantic vacuity of σ, with respect to a situation w, is the ratio between the cardinality n of the set of all the situations in E, necessarily including w, that support σ, and the number s^{l} of possible situations in E.

In Vac, n is a positive integer that satisfies the condition $1 \leq n > s^{l}$. The number of situations supporting $\sigma \in \Sigma^{x}$ determine the specific value of n. Is there a systematic method for generating a "continuum" of progressively weaker synthetic messages in E whilst keeping l constant? A simple solution is provided by the introduction of semiduals.

Semiduals are duals in which only the connectives $\{\wedge, \vee\}$ have been inverted but atomic messages have not been replaced by their complement, whilst parentheses can be introduced to avoid ambiguities. A message of length l contains $l-1$ connectives. We start with the assumption that all $l-1$ connectives are conjunctions, that is all messages are strongly synthetic. By replacing all conjunctions with $l-1$ disjunctions we obtain the set of most weakly synthetic messages in E, whose $\vartheta = (s^{l}-1)/s^{l}$. We can then proceed to replace $l-2$ conjunctions, $l-3$ conjunctions and so forth, until $l-n$ disjunctions $= 1$. The set with $l-n = 1$ disjunction is the set of true, synthetic messages in E with the lowest number of supporting situations and hence the lowest $\vartheta > 0$ in the model. In E, this means that we can construct five classes, whose members are true and consistent with, respectively, 63, 35, 15, 7, or 3 situations. These are the corresponding values of n.

Vac and the previous method of semantic weakening allow us to partition Σ into $l-1$ disjoint classes Vac $= \{Vac_{1} \ldots Vac_{l-1}\}$, and map each class to its corresponding degree of vacuity ϑ. Note that the value of ϑ for the class of most weakly synthetic messages in E is

always $(s^l-1)/s^l$, whilst the middle values are found according to the formula $((s^l/2)-1)/s^l$, $((s^l/2^2)-1)/s^l$, … $((s^l/2^n)-1)/s^l$, where $(s^l/2^n) - 1 = 3$, i.e. the smallest number of situations, including w, consistent with σ_i when $\vartheta(\sigma_i) > 0$ (see Table 4 for an application to the model E).

6. Degrees of informativeness

We are now ready to calculate the *degree of informativeness function* of σ. The complement of the squared value of $\vartheta(\sigma)$ with respect to the maximum degree of informativeness provides an accurate measure of the informativeness of σ (see Table 5 for the graphics, ι values are mapped onto the y-axis):

DI $$\iota(\sigma) = 1 - \vartheta^2(\sigma)$$

What are the motivations for DI? If possible, the equation should satisfy the following 6 constraints, derived from the five necessary conditions for a satisfactory metrics M.1-M.5:

E.1 $$(\vartheta(\sigma) = 0) \rightarrow (\iota(\sigma) = 1)$$

E.1 follows immediately from M.1.

E.2 $$\int_a^b \boldsymbol{i}(\boldsymbol{s})\, dx \text{ is a proper integral}$$

E.2 follows from the fact that the function $f(\sigma)$ is bounded on the interval [0, 1]. In the next section, we shall see that E.2 simplifies the calculation of the quantity of informative, vacuous or inaccurate content in σ.

E.3 $$(\vartheta(\sigma) = (+1 \vee -1)) \rightarrow (\iota(\sigma) = 0)$$

E.3 follows immediately from M.2 and M.3.

E.4 $$((0 < \vartheta(\sigma) < +1) \vee (0 > \vartheta(\sigma) > -1)) \rightarrow 0 < \iota(\sigma) > 1$$

E.4 follows immediately from M.4 and M.5.

E.5 $\qquad$ a small variation in $\vartheta(\sigma)$ results in a substantial variation in $\iota(\sigma)$.

E.5 is meant to satisfy the requirement according to which the lower $\iota(\sigma)$ is, the smaller is the possible increase in the relative amount of vacuity or inaccuracy carried by σ. E.5 will become clearer in the next section, once the two concepts of quantity of vacuity and inaccuracy are introduced. Here, it is possible to rely on the intuitive view that, moving from a σ with $\vartheta = 0$ towards a σ with either $\vartheta = +1$ or $\vartheta = -1$, the first steps can be expected to bring a comparatively greater loss of informativeness (a greater increase in vacuity or inaccuracy) than the following ones. More generally, this means endorsing the view that an information system (a) is not brittle, like a classic logic system (the presence of an inconsistency in the former is not as destructive as in the latter), and (b) does not have a progressive degree of fault tolerance (in the case of examination assessment techniques or in the context of assessment of moral responsibility, for example, *ceteris paribus*, the errors are usually evaluated more and more severely, the second having a comparatively more negative impact than the first and so forth). It is rather described as having an "inverted" degree of fault tolerance: faults are decreasingly less impairing, the first being more damaging than the second and so forth (this holds true, for example, in the case of experiments error analysis, or in cases of assessment of moral trust and faithfulness).

E.6   the *marginal information function* MI is a linear function.

E.6 is justified by the requirement that, a priori, all atomic messages ought to be assigned the same potential degree of informativeness and therefore, although E.5 indicates that the graph of the model has a variable gradient, the rate at which $\iota(\sigma)$ changes with respect to change in $\vartheta(\sigma)$ should be assumed to be uniform, continuous and linear.

DI satisfies E.1 by adopting the standard convention according to which the absolute maximum value $\iota(\sigma) = +1$ occurs at $\vartheta(\sigma) = 0$, by analogy with the range of values of $\text{CONT}(\sigma)$ and $p(\sigma)$. It satisfies E.2 by calculating the degree of informativeness of σ in terms of the complement of the value of $\vartheta(\sigma)$, not its reciprocal. We saw in section 3 that this solution has

the further advantages of being mathematically simpler and in agreement with TWSI's approach. Finally, DI satisfies E.3-E.6 by referring to the *squared* value of $\vartheta(\sigma)$. From DI, the derivative of $\iota(\sigma)$ is:

MI $\qquad\qquad \iota'(\sigma) = d/dx\ 1 - \vartheta^2(\sigma) = -2\vartheta(\sigma)$

MI means that the total degree of informativeness of σ changes at a rate of -2ϑ. The squared value of ϑ makes the function continuous and differentiable and provides the most satisfactory solution to model DI.[25]

7. Quantities of vacuity and semantic information

Now that it is possible to calculate the *degree* of informativeness of σ, the next step is to calculate on its basis the relative *quantity* of semantic information in σ. To this purpose, the absolute maximum quantity of information MQI and its unity must first be defined.

The value of MQI is inferred from DI, by calculating the definite integral of the function $\iota(\sigma)$ on the interval [0, 1]:

(MQI) $\quad \int_0^1 \iota(\sigma)\, dx = Log_2 \frac{2}{3} = Log_2 2 - Log_2 3 = 1\, bit - 1\, trit = 1\, sbit$

The term *sbit* (*semantic bit*) indicates the unit of semantic information in TSSI. One sbit corresponds to the maximum quantity of semantic information, concerning the fixed situation w, that can be conveyed by an infon σ with $\vartheta = 0$. It is linearly equivalent to Log_2 2/3.[26]

$\vartheta^*(\sigma)$, the quantity of vacuous information in σ, can now be obtained from [MQI] by calculating the ratio between the definite integral of the function $\iota(\sigma)$ on the interval [0, ϑ] and 1 sbit:

QV $\qquad\qquad J*(s) = \dfrac{\displaystyle\int_0^J i(s)\, dx}{1\ sbit}$

Finally, the quantity of semantic information in σ, $\iota^*(\sigma)$, can be obtained from QV, by calculating the complement of ϑ^* with respect to 1 sbit:

QI $\qquad\qquad\qquad \iota^*(\sigma) = 1 \text{ sbit} - \vartheta^*(\sigma)$

As required by the model constraints, when $\vartheta(\sigma) = 0$, it follows that $\vartheta^*(\sigma) = 0$ and $\iota^*(\sigma) = 1$ sbit (σ is a fully accurate, precise and contingent truth), and when $\vartheta(\sigma) = 1$, it follows that $\vartheta^*(\sigma) = 1$ sbit and $\iota^*(\sigma) = 0$ (σ is a tautology). Note that, following QI, the quantity of informative content in σ is established by reference to its truth-oriented properties and is not immediately identified with all information that is nomically or analytically nested in σ, as in Dretske [1981], yet the two approaches are perfectly compatible.

8. The solution of the Bar-Hillel-Carnap Paradox

We are now in a position to evaluate TSSI's solution of BCP. Suppose a deflationary argument is offered, phrased as follows. BCP owes its apparently counterintuitive nature to a conceptual confusion between CONT(σ) in 1 and $\iota^*(\sigma)$ in QI. Once the two concepts are distinguished with sufficient clarity, the alleged paradox vanishes. On the one hand, CONT(σ) refers to the *quantity of semantic information* that can be attributed to σ a priori, on the basis of its probability distributions and independently of the state in which the system under analysis actually is (context of total ignorance). On the other hand, $\iota^*(\sigma)$ refers to the *quantity of semantic information* that can be attributed to σ still a priori but in a context which is presupposed to be of "localised omniscience" in the game-theoretical sense of perfect and complete information about the system, on the basis of σ' alethic value and its degree of discrepancy, relative to a fixed state w of the system under analysis. So R.1 and R.2 really embody two different and hence compatible explications of the pre-theoretical idea of "quantity of semantic information": R.1 refers to the relation between CONT(σ) and p(σ),

whereas R.2 refers to the relation between $\iota^*(\sigma)$ and $\iota(\sigma)$. Since the two requisita are not really in conflict, there is no paradox to be solved.

Unfortunately, the conclusion cannot be granted. The previous analysis has shown that $\text{CONT}(\sigma) \neq \iota^*(\sigma)$, but this inequality is insufficient to explain the counterintuitive definition of the former in TWSI. What does follow from the deflationary argument is that the paradox can no longer be satisfactorily explicated merely in terms of a semantic conflict between R.1 and R.2. But the counterintuitive nature of $\text{CONT}(\sigma)$ is actually increased by the fact that it is now obvious that $\text{CONT}(\sigma)$ does not provide an indication of the amount of *informativeness* of σ. What does $\text{CONT}(\sigma)$ really purport to indicate then? Despite the fact that $\text{CONT}(\sigma) \neq \iota^*(\sigma)$, there is a clear conceptual connection between the two measures: they both attempt to provide, from different perspectives, a quantitative evaluation of the "information-richness" of σ. This is why it is difficult to reconcile $\text{CONT}(\sigma)$ with a sound understanding of what the quantity of semantic information conveyed by σ is, without any further proviso. There is no uncontroversial sense in which a contradiction or a lie can be richer in information than a true proposition, and actually TSSI shows in what sense exactly the opposite can be proved. Abandoning $\text{CONT}(\sigma)$ as a useless notion, marred by paradoxical implications, would be too hasty. Transforming it into a mere definitional convention or adopting *ad hoc* solutions, though viable alternatives, would not only be in conflict with Carnap's and Bar-Hillel's original interpretation of their semantic theory, but would also make it substantially less interesting. Given the usefulness of TWSI, $\text{CONT}(\sigma)$ should probably be salvaged, if possible. In TWSI, $\text{CONT}(\sigma)$ is meant to indicate nothing less than the quantity of semantic information carried or conveyed by σ_m, which may be greater than the quantity of semantic information carried or conveyed by σ_n (Bar-Hillel [1964], 222 and 299). It is exactly the unqualified boldness of this general claim that is in need of a more

circumspect formulation. In order to avoid BCP, two modifications in the understanding of $\text{CONT}(\sigma)$ are in order: a clarification of what $\text{CONT}(\sigma)$ is really a measure of, and a constraint on its applicability.

First, $\text{CONT}(\sigma)$ does not indicate the quantity of *semantic information* but, more precisely, the quantity of *data* in σ. By concentrating on the degree of systemic consistency and then the logical probability of sets of infons, $\text{CONT}(\sigma)$ deals only with *completely uninterpreted information*, that is data, which are not carried by, but actually constitute σ, as syntactically well-formed combinations of interpretable signs or signals. Therefore, between TWSI and TSSI, it is the latter that, working on *alethically-interpreted data*, comes closer to providing a quantitative indication of the information-richness or poverty of σ. The former evaluates the information-richness of messages only insofar as their implementation is logically possible.

Secondly, because TWSI does not deal with semantic information but only with data and their possible combinations a priori, the general equation $\text{CONT}(\sigma) = 1 - p(\sigma)$ in R.1 cannot be assumed to have a precise value independently of its frame of reference. This means that the equation is really meaningful only once it is properly constrained by the following three systemic conditions:

S.1) unambiguous individuation and explicit description of the data system (in our model a universe consisting of 3 constants, each qualifiable by two properties, affirmed or negated);

S.2) generation of a fully normalised description of the data system, as the set of all the mutually exclusive messages necessary and sufficient to describe in full all the possible states in which the data system can be;

S.3) attribution of a probability p to each state-description σ_i—either in terms of uniform distribution, or according to any bias for which there may be evidence in the system—such that the following two standard conditions are satisfied:

$$0 < p(\boldsymbol{s}_i) < 1, \text{for all i}$$

$$\sum_{i=p}^{n} p(\boldsymbol{s}_i) = 1$$

It is only at this point that a precise meaning can be attached to CONT(σ). The complement of the value of p(σ) with respect to the maximum degree of probability can be used to indicate the quantity of data in σ only because of the connection between S.1 and S.2, whilst p can only be assigned a value in the open interval (0,1) because of S.3. It is not that we are unable to analyse the probability of σ absolutely, but rather that the latter is inappropriate for the task in question. The probability of σ can be correctly interpreted as a measure of the quantity of interpretable data in σ only when σ is completely *uninterpreted* and *implementable*, that is only when we are dealing with syntactically well-formed strings of symbols or signals which are not already known to be either necessarily true or false a priori. These constraints make it possible to couple[27] probability of σ and quantity of data in σ sufficiently tightly that the former can provide a reliable indication of the latter. Outside these constraints, the two measures may not be significantly related or even lead to paradoxical conclusions, as BCP shows.

The two modifications in the interpretation of CONT(σ) lead to a re-assessment of the meaning of the standard view concerning the relation between the information-richness of σ and its likelihood. To develop a clear understanding of semantic information we need to move from likelihood (TWSI) to lkeness (TSSI), as it were. When we say that the less likely σ is the more informative it may be assumed to be, unless we are making some psychologistic remark about the subjective expectations of a user, we are referring to the higher or lower degree of discrepancy of σ with respect to one or more *w*. The less vacuous σ

is the fewer possible worlds support it, and the more informative σ becomes with respect to the fixed w that acts as a benchmark. Of course, this direct relation can also be analysed in terms of probability, since the latter can provide an interpretation for the concept of vacuity. But the two concepts do not overlap and they diverge when the σ in question is no longer uninterpreted. Information is an actual possibility that is inconsistent with at least one but not all other possibilities. A contradiction is not information-rich because it is not a possibility, and a tautology is not information-rich because it does not exclude any possibility. In TSSI, they are both limit instances of "uninformation" (lack of both positive information and negative misinformation).[28]

9. Conclusion: summary of results and future developments

In this paper, a quantitative theory of strongly semantic information (TSSI) has been shown to be possible on the basis of a calculus based on truth-values and degrees of discrepancies with respect to a given situation, rather than probability distributions. The main hypothesis supported has been that semantic information encapsulates truth, and hence that false information fails to qualify as information at all. The expression "false information" is to be equated to expressions such as "false policeman" (not a policeman at all) or "false passage" (not a passage at all), not to "false leg" (still a leg, though artificial). The main result of the development of TSSI has been the solution of the semantic paradox affecting the classic quantitative theory of semantic information, according to which a contradiction contains the highest quantity of information. In the course of the analysis, the paper has provided a review of the requirements for any quantitative theory of semantic information, of the criteria of semantic information equivalence, of the concepts of degrees of strongly (i.e. truth-based) semantic inaccuracy, vacuity and informativeness; and of the concepts of quantities of

strongly semantic vacuity and information. Three results of conceptual interest, based on TSSI, which have been left to a second stage of the research are:

1) the extension of the quantitative analysis to the semantic concepts of quantity of misinformation (ordinarily called "false information") and degree of disinformation, as foreshadowed in D.3;

2) the generalisation of the results, obtained by TSSI in connection with the formal logical setting represented by the finite model system E, to the more general context represented by (regions of) the infosphere, where the latter is understood as the universal system of information and is modelled by means of *Formal Methods*[29] (popular model-based formalisms consistent with the present treatment are given by the specifications languages *Z* and *VDM*, see Woodcock and Davies [1996] and Jones [1986]); and

3) the analysis of the standard definition of knowledge as true justified belief, in light of a "continuum" hypothesis that knowledge encapsulates truth because it encapsulates semantic information.[30]

References

Aisbett J. and Gibbon G. 1999, "A practical measure of the information in a logical theory", *Journal of Experimental and Theoretical Artificial Intelligence* (11.2), pp. 201-218.

Bar-Hillel Y. 1964, *Language and Information* (Reading, Mass.; London: Addison-Wesley).

Bar-Hillel Y. and Carnap R. 1953, "An Outline of a Theory of Semantic Information", rep. in Bar-Hillel [1964], pp. 221-274, page references are to this edition.

Barwise J. and Perry J. 1983, *Situations and Attitudes* (Cambridge, Ma.: MIT Press).

Cherrry C. 1978, *On Human Communication* 3rd ed. (Cambridge, Ma: MIT Press).

Devlin K. 1991, *Logic and Information* (Cambridge: Cambridge University Press).

Dretske F. 1981, *Knowledge and the Flow of Information* (Cambridge, Ma: MIT Press, rep. Stanford: CSLI, 1999).

Floridi L. 1999, *Philosophy and Computing* (London – New York: Routledge).

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

Hanson W. H. 1980, "First-Degree Entailments and Information", *Notre Dame Journal of Formal Logic* (21.4), 659-671.

Heck A. and Murtagh F. (eds.) 1993, *Intelligent Information Retrieval: the Case of Astronomy and Related Space Sciences* (Dordrecht - London: Kluwer Academic).

Hintikka J. and Suppes P. (eds.) 1970, *Information and Inference* (Dordrecht: Riedel).

Hockett C. F. 1952, "An Approach to the Quantification of Semantic Noise", *Philosophy of Science* (19.4), pp. 257-260.

Jamison D. 1970, "Bayesian Information Usage", in Hintikka and Suppes [1970], pp. 28-57.

Jeffrey, R. C. 1990, *The Logic of Decision*, 2nd ed. (Chicago: University of Chicago Press).

Jones C. B. 1986, *Systematic Software Development Using VDM* (London: Prentice-Hall International).

Kemeny J. 1953, "A logical measure function" *Journal of Symbolic Logic* (18), pp. 289-308.

Levi I. 1967, "Information and Inference", *Synthese* (17), pp. 369-391.

Lozinskii E. 1994, "Information and evidence in logic systems", *Journal of Experimental and Theoretical Artificial Intelligence* (6), pp. 163-193.

Mingers J. 1997, "The Nature of Information and its Relationship to Meaning" in Winder [1997], pp. 73-84.

Popper K. R. 1935, *Logik der Forschung* (Vienna: Springer) trans. *The Logic of Scientific Discovery* (London: Hutchinson, 1959).

Popper K. R. 1962, *Conjectures and Refutations* (London: Routledge).

Reza F. M. 1994, *An Introduction to Information Theory* (New York: Dover, orig. 1961).

Shannon C. E. 1948, "A Mathematical Theory of Communication", *Bell System Tech. J.* (27), pp. 379-423, 623-656.

Shannon C. E. 1993, *Collected Papers* (Los Alamos, Ca: IEEE Computer Society Press).

Shannon C. E. and Weaver W. 1949, *The Mathematical Theory of Communication* (Urbana, Ill.: University of Illinois Press).

Smokler H. 1966, "Informational Content: A Problem of Definition", *The Journal of Philosophy* (63.8), pp. 201-211.

Sneed D. J. 1967, "Entropy, Information and Decision", *Synthese* (17), pp. 392-407.

Sorensen R. 1988, *Blindspots* (Oxford: Clarendon Press).

Szaniawski K. 1967, "The Value of Perfect Information", *Synthese* (17), pp. 408-424, now in Szaniawski [1998].

Szaniawski K. 1974, "Two Concepts of Information", *Theory and Decision* (5), pp. 9-21, now in Szaniawski [1998].

Szaniawski K. 1984, "On Defining Information", now in Szaniawski [1998].

Szaniawski K. 1998, *On Science, Inference, Information and Decision Making, Selected Essays in the Philosophy of Science*, ed. by A. Chmielewski and J. Wolenski (Dordrecht: Kluwer).

Taylor J. R. 1997, *An Introduction to Error Analysis: The Study of Uncertainty in Physical Measurements*, 2nd ed. (Mill Valley, CA: University Science Books).

Taylor K. A. 1987, "Belief, Information and Semantic Content: A Naturalist's Lament", *Synthese* (71), pp. 97-124.

Van der Lubbe J. C. A. 1997, *Information Theory* (Cambridge: Cambridge U.P., orig. 1988).

Williamson T. 1994, *Vagueness* (London: Routledge).

Winder R. L., Probert S. K. and Beeson I. A. 1997, *Philosophical Aspects of Information Systems* (London: Taylor and Francis).

Woodcock J. C. P. and Davies J. 1996, *Using Z: Specification, Refinement and Proof* (London: Prentice-Hall International).

Notes

1 This is short for "semantic information content". This paper follows the common practice of using the two expressions interchangeably.

2 The first version of the paper was read before the Symposium on Applications of Communication Theory in 1952. A more detailed and systematic treatment appeared then in Technical Report No. 247 of the Research Laboratory of Electronics, M.I.T., 1953. A slightly revised version was published in *The British Journal for the Philosophy of Science* in 1954. The chapter in Bar-Hillel [1964] is from the 1953 version, which is actually dated 1952 in the opening footnote.

3 Cherry [1978], for example, contains one of the clearest and most informative summaries of TWSI, but no reference to the paradox.

4 Beginning with Bar-Hillel and Carnap [1953]. Bar-Hillel's and Carnap's analysis, discussed in the text, is further developed in Kemeny [1953], Smokler [1966] and Hintikka and Suppes [1970].

5 This is also known as Mathematical Theory of Information, or Communication Theory, Information Theory and Mathematical Information Theory. I have opted for STST in order to avoid any possible confusion. One of the best conceptual introductions to STST is still Dretske [1981], but see also Cherry [1978]. More technical presentations can be found in Reza [1994] and Van der Lubbe [1997].

6 For a defence of the view that false information is not an inferior kind of information but not information at all see Dretske [1981] and Grice's "Logic and Conversation" in Grice [1989], especially p. 371 of the "Retrospective Epilogue".

7 The classic references are Shannon [1948], Shannon and Weaver [1949], Shannon [1993], but see also footnote 5 above. I shall avoid in this context any reference to "doubt"

(uncertainty of the signal) and "surprise value" (statistical rarity of the signal), for these psychologistic metaphors are not very helpful.

[8] Popper [1935] was one of the first to suggest that the amount of semantic information in a statement can be analysed in terms of the number of alternative possibilities excluded by that statement, see also Popper [1962]. For a *positive* account, in terms of the set of all state-descriptions entailed by σ, see Hanson [1980]. Both accounts equally lead to the formulation of BCP.

[9] Following Hockett [1952], the choice of the model is suggested only by syntactic elegance and simplicity, and results can be adapted to cases where some redundancy also occurs.

[10] The model is useful precisely in order to have a discrete, one-to-one correspondence between formulae–like truth–makers and semiotic truth–bearers. This idealisation guarantees information completeness. The question of whether a more coarse–grained model of truth–makers is metaphysically preferable will not be addressed here. Strictly speaking, one needs to recall that, as in error analysis theory (Taylor [1997]), infons (e.g. measures) are never compared to "absolute realities" (e.g. absolutely true values), but always and only to other infons that are assumed or known to be true, i.e. states of the world that are transformed into evidence by the information process, at least momentarily, that is until further critical revision (Levi [1967]).

[11] More formally: [1] $\text{CONT}(\top) = \text{MIN} = \forall\sigma_n\forall\sigma_i\,(((\models\sigma_n \wedge \not\models\sigma_i) \to (\text{CONT}(\sigma_n) < \text{CONT}(\sigma_i))) \wedge ((\models\sigma_n \wedge \models\sigma_i) \to (\text{CONT}(\sigma_n) = \text{CONT}(\sigma_i))))$

[12] More formally: [BCP] $\text{CONT}(\bot) = \forall\sigma_n\forall\sigma_i\,(((\sigma_n\models \wedge \sigma_i\not\models) \to (\text{CONT}(\sigma_n) > \text{CONT}(\sigma_i))) \wedge ((\sigma_n\models \wedge \sigma_i\models) \to (\text{CONT}(\sigma_n) = \text{CONT}(\sigma_i))))$.

13 "Normalisation" refers here to the process followed to obtain a database design that allows for efficient access and storage of data and reduces data redundancy and the chances of data becoming inconsistent.

14 See Sneed [1967] for an early, critical analysis based on Jeffrey [1967], now Jeffrey [1990]. Szaniawski [1967], based on a game-theoretic approach, and Szaniawski [1974], both now in Szaniawski [1998], is more optimistic. Smokler [1966] defends a moderate approach, which, however, requires that every individual has, at a certain time, only a finite and consistent set of true/false beliefs, and that these can be expressed in the language of the classic logic of predicates. These seem unrealistic requirements, especially since Smokler accepts both (i) that beliefs are best characterised as propositions and (ii) that the latter are non-linguistic entities. As a counterexample, it suffices to note that any individual who knows arithmetic can both believe an infinite number of true propositions and hold some contradictory beliefs. Jamison [1970] goes as far as to argue that a pragmatic theory of informational quantities is the most fundamental of all approaches, and presents a sort of inverted BCP by suggesting that (p.29) "an undesirable feature of RL [i.e. what has been defined in this article as TWSI] is that in it logical truths carry no information", so that mathematical equations, insofar as they can be interpreted as tautologies, would be utterly non-informative.

15 "The name [infon] suggests a parallel with the fundamental particles of physics, the electrons, protons, neutrons, photons and so forth." (Devlin [1991], p. 37) Is there a conceptualisation of "information" as "a theoretical commodity that we can work with, analogous to (say) the numbers that the number-theorist works with, or the points, lines and planes the geometer works with"? (Devlin [1991], p. 18) The question is answered in Dretske [1981], who provides a definition of information as an objective commodity. However,

Taylor [1987] is justified in arguing that Dretske's "objectivism" is not easily reconcilable with his pragmatic approach, whereby information is also defined pragmatically, with reference to a user S.

[16] The context is ideal because it presupposes a state of *perfect* and *complete* information about the system, in the game-theoretic sense of the expressions. We shall see in section 9 that this is a further difference between TWSI and TSSI, which presupposes a state of imperfect and incomplete information with respect to the states of the system. Here it is possible to anticipate that presupposing some "localised omniscience" does not affect the value of the approach, since TSSI attempts to discover what the quantity of semantic information in σ is where σ's alethic value is known, not the average amount of semantic information any σ may have a priori.

[17] Szaniawski [1984] does not refer explicitly to any result in situation logic, but develops an approach similar to Devlin [1991]. A limitation of his analysis is that he does not acknowledge the fact that the change in perspective entails an inversion in the relation between information and truth, which no longer supervenes on, but now constitutes the former. His definition of semantic information as "(p. 231) potential information about X that has been 'singled out' in some way: observed, asserted, etc. Semantic information may be said to be true if it points to the actual state of the world" is not satisfactory; not so much because it is circular (in the article the circularity is avoided at the cost of some vacuity, by means of a reference to unspecified ways in which σ is "relevant" and "points" to the actual state in which X is) as because it does not clarify what difference is made by the "singling out" procedure. In his article, it becomes clear that "singling out" p is attributing a positive alethic value to p as a description correctly pointing to the actual state in which X is.

[18] This is a technical term in error analysis terminology. When a measurement or a result is compared to another that is assumed or known to be more reliable, the difference between the two is defined as the experimental discrepancy.

[19] This is another technical term in error analysis. Deviations are "experimental uncertainties" and are commonly called "errors". More precisely, when a set of measurements is made of a physical quantity, the difference between each measurement and the average (mean) of the entire set is called the deviation of the measurement from the mean.

[20] Although ϑ can in principle take any value in the range of real numbers $[-1.0, +1.0]$, thus giving rise to a continuous function (see Table 5), this does not mean that, given a specific model, an infinite number of interpretations of ϑ is effectively available. More realistically, a model always sets up a finite and discrete range of values for ϑ, approximate to n decimals. Following standard practice, in what follows the function is defined for all real numbers in the specified range, and discontinuities are associated only with truly dramatic behaviour on the part of the model.

[21] Note that in this context vacuity does not refer to typical problems raised by borderline cases or the sorites paradox. In the model, each situation w is precise and each infon σ sharply divides the world into those situations to which σ applies and those to which it does not. In this context, vacuity is a matter of semantic uncertainty, not in the sense that infons with a certain degree of vagueness are infons that apply to situations to varying degrees, hence generating indecision about borderline cases, but in the sense that infons with a certain degree of vagueness are infons that apply to a varying number of situations, hence generating indecision with respect to which w is the case. Infons do not increase gradually in their degrees of truth, as in fuzzy logic, but in their degree of conformity to w (lack of

discrepancy). This approach appears to be consistent with the epistemic view according to which vagueness is a kind of ignorance (Sorensen [1988], Williamson [1994]).

[22] "Inaccuracy" here does not have exactly the same technical meaning as in error analysis theory, where a measurement is said to have high precision if it contains a relatively small indeterminate error, and is said to have high accuracy if it contains relatively small indeterminate error and relatively small determinate error.

[23] Errors are not to be understood as blunders. Consistently with current practice in error analysis theory, the term "error" is used here as a synonym for "logical uncertainty" (see above the comment on "uncertainty" as "occurrence/elimination of possibilities"). Hence, the "errors" in Inac are independent of the user, precisely characterizable and can be assumed to be normally distributed, unless the system is biased.

[24] Table 1 is laid out in such a way that the fourth column contains the complements of the first column upside down (e.g. w_{49} is the complement of w_{16}), and the third column the complements of the second.

[25] Why? Because in $\iota(\sigma) = 1 - \vartheta(\sigma)^n$, n could be an odd integer, 1, or an even integer greater than 2, but none of these alternatives is fully satisfactory. If n is an odd integer, the model satisfies only E.2, and introducing the absolute value of ϑ still leaves E6 unsatisfied. If $n = 1$, the equation satisfies E.1, but in order to satisfy E.2 and E.3 we need to calculate the complement of the absolute value of $\vartheta(\sigma)$. However, the new formula $1 - |\vartheta(\sigma)|$ still fails to satisfy [E.6]. If n is any even integer greater than 2, the equation not only represents merely a more complicated extension of the simpler solution adopted in DI, it also fails to satisfy E.6.

[26] A trit is one base-3 digit and represents the amount of information conveyed by a selection among one of three equally likely outcomes. It is linearly equivalent to $\text{Log}_2 3$ bits.