Do Logical Truths Carry Information?

ABSTRACT

The paper deals with the question whether logical truth carry information. On the one hand it seems that we gain new information by drawing inferences or arriving at some theorems. On the other hand the formal accounts of information and information content which are most widely known today say that logical truth carry no information at all. The latter is shown by considering these accounts. Then several ways to deal with the dilemma are distinguished, especially syntactic and ontological solutions. A version of a syntactical solution is favoured.

KEYWORDS
Logical truth, information content, logical form, algorithmic information, depth information

Introduction

This paper deals with a special problem within the field of explicating the/a concept of information: the problem of the information content of logical truth. Getting to know some logical truth seems to involve acquiring some information. This applies especially to drawing consequences. In drawing a consequence we get information about what was entailed or implied by what we already believed. Getting to know a consequence relation between some beliefs or sentences seems to get information.

Nevertheless the standard explications of information content are not able to deal with the problem of the information content of logical truth. I will highlight this problem in the different approaches. I then distinguish several strategies to deal with the problem, i.e. strategies to assign logical truth either information content or some other quality accounting for the gain in knowledge upon acquiring them. Some ways to look for a solution to the problem are hinted at, but none has been fully developed so far.

§1 The Problem in the Syntactic Approach

Rational students should engage only in courses where they can learn something. Now, unfortunately, it seems that you can learn nothing in a logic course, if learning something means acquiring some information, since the information content of logical truths – seen in the light of the standard approaches to measuring or defining information content – is: *nothing*!

Let us consider the mathematical theory of communication as developed by Claude Shannon (1949) first. The average information of a source is defined given some measure of the probability that some symbol out of a set of symbols occurs and the uncertainty with which that symbol occurs given the possible strings of symbols made out of the symbols in that set. Starting from some requirements on the notion of information content (like information being additive and that information decreases uncertainty) Shannon uses a logarithmic measure of the uncertainty of a symbol, a binary coding and a probability measure to derive his famous formula:

(1) $H = - \Sigma_{i=1}(P_i \log P_i)$

This refers to the source as a whole. Applied to a single signal we can say:

(2) $I(\alpha) = \log(1/p(\alpha))$

The amount of information in a single symbol α (whether letters in a word or sentences taken as the single units in talking) is the logarithm of the reverse of its probability.

Logical truths are not random. They can be completely expected, there are no alternatives to them. Their probability is 1. This means that, in the syntactic approach, given the definition of information content "I(x)", we get for a logical truth α:

(3) $I(\alpha)= \log(1/p(\alpha)) = \log(1/1) = \log 1 = 0$

Logical truth carry no information at all. You learn nothing from them.

§2 The Problem in the Semantic Approach

Carnap and Bar-Hillel (1952) developed a semantic theory of information content within the possible worlds framework. Their analysis from the very beginning concerns sentences not individual letters or symbols. As usual, one might identify what a sentence says with the set of possible worlds in which the sentence is true. The information content of a sentence might be taken as the set of worlds excluded by this sentence being true, since in this way we keep the intuition that information content is related to surprise that what a symbol says is the case. So Carnap and Bar-Hillel develop two explications of semantic content. The one starts with the idea just mentioned and gives a more semantic measure of information content, since the range of worlds excluded by a sentence is statically associated with that sentence. It does not change with our knowledge which world is the actual world. A measure *cont* can be gained by counting the excluded worlds or by employing an a priori probability measure which assigns all worlds the same probability. Let m be such a measure, m(α) is the probability of a sentence α. Then we can define cont:

(1) $\text{cont}(\alpha) = 1 - m(\alpha)$

Logical truths are true in all possible worlds. The set of the worlds excluded by their truth is ∅, i.e. given the explication "cont( )" of information content in the possible worlds approach:

(2) cont(α) = ∅ (collecting the excluded worlds)

or cont(α) = 1 – m(α) = 1 – 1 = 0

Given a probability measure on worlds the information content of a logical truth α is the number of (the sum of the probability of) the worlds in ∅, i.e. 0, or the reverse of the probability of α, i.e., once again 0. Given cont a logical truth carries no semantic information at all although logical truth are, given Carnap`s semantic model, true because of their meaning.

Considering that we have quite different intuitions with respect to information content Carnap and Bar-Hillel provide a second explication of semantic content in terms of probability (given any probability distribution on the set of possible worlds) and a logarithmic measure. This second measure is more epistemic than semantic, since the probability distribution we choose might reflect our world knowledge. With this measure they derive a semantic analogue to Shannon`s formula:

(3) inf(α) = - log(m(α))

Repeating the calculation from the last paragraph we get:

(4) inf(α) = - log(m(α)) = - log 1 = 0

Once again you learn nothing from logical truth.

Luciano Floridi (forthcoming) developed the semantic approach into a theory of "strongly semantic information". His starting point stresses one of the contra-intuitive consequence of the original semantic approach: that contradictions have the maximum information value. This holds true in the Carnap/Bar-Hillel framework since contradictions exclude all possible worlds; their range being ∅ means that the reverse of their range is the totality of possible worlds. Their probability is zero. And the reverse of their probability is therefore maximal. If α is a contradiction:

(5) cont(α) = 1 – m(α) = 1 – 0 = 1

(6) inf(α) = log(1/m(α)) = log (1/0) ≈ log ∞ = ∞

Floridi calls this „the Bar-Hillel/Carnap paradox", since, intuitively, we would say that somebody who utters a contradiction has said nothing at all, has conveyed no information at all. He develops a theory in which we not only consider the truth value of a sentence but also the amount of its deviation (in degrees) from the actual world (like "there are eight dogs" deviates more from a situation with two dogs than “there are six dogs”, although both sentences are false). Given his account of discrepancy of a sentence from the actual world, he can derive that the discrepancy of contradictions is maximal, which means that their information content is zero. So he can in fact solve the problem of the supposedly informative contradictions. As one condition in the development of the appropriate information content function, however, he explictly lays down the condition that if α is a tautology it is assigned the maximum degree of discrepancy. That makes it a part of his framework that logical truths carry no information. So even this elaborated semantic approach refuses to give us information from logical truths.

§3 The Problem in Dretske´s Approach

Fred Dretske (1981) developed an account of information that preserves the main ideas of the syntactic approach and tries to combine it with an externalist account of semantic information content. It takes information as being out there in the world. Meanings might be partly in the head but information is not. Information flows because of the causal connections between some object s being F and another object b is G. Dretske does not consider average amounts of information associated with some symbol but an absolute content given a framework of natural laws and the circumstances of the situation. So s being F *carries the information* that b is G if the conditional probability of b being G given s being F is 1. (A conditional probability of less than 1 will not do, because of some criteria on information flow like his famous “Xerox-Principle”.) *Knowledge* is *defined* as the belief that s is F caused by the information that s is F, *given* some natural laws. The natural laws and so, of course, the laws of logic belong to the framework within information flow is recognised. What belongs to the framework cannot carry information itself. Even natural laws, as given in *all* relevant contexts, “have an informational measure of zero” (ibid p.264). Logical truth do not cause anyway. So in Dretske´s externalistic approach to information the problem of non-contingent (logical) truth is even more pressing. Since you have the framework already you can learn nothing from a logical truth.

§4 How to Solve the Problem?

There might be different types of solution:

a) logical truths carry no information in the sense explained,
   but are nevertheless of interest because of some other quality.

This type of solution would leave information theory at it is but supplements it with a theory of what happens in recognising logical truths besides information flow as explained by the standard accounts.

b) information is analysed so as to be able to distinguish between some logical truths.

A kind of syntactic approach can be of type (a), an ontological approach of type (b).

4.1 A Syntactic Solution

Within a semantic approach some *syntactic* features can be given a role:

The logical truths

(1) $p \supset p$

and (2) $(\forall x)(x=x)$

differ syntactically. Carnap´s concept of *intensional isomorphy* (Carnap 1955) introduces some syntactic features into an account of meaning. Two sentences are intensional isomorphic if one can be transformed into the other substituting step by step expressions of the same syntactic category for each other. Since (2) contains expression of the syntactic type individual (variable) it cannot be transformed into (1). We can introduce a concept of meaning that not only requires logical equivalence but also requires that two logical true sentences can only have the same meaning if at their deepest level of logical form they share one logical form (Bremer 1993:295-96). So (1) and (2) *differ* in *meaning*. We care about differences in meaning so that would be an account why we care about different logical truth. Each logical truth tells us that some individual sentence (i.e. a sentence with a meaning that distinguishes it from all other sentences) is a logical truth. In recognising a consequence relation we see a connection between meanings that we did not see before.

Another version of such a syntactic solution could be developed within a computational theory of mind which refers to mental representations (maybe some *language of thought* symbols). Within such a computational theory of mind mental representations have their semantic features and their (psychological) role because of their syntactic features, since only these configurations enter into causal connections (cf. Fodor 1987, 1994).

The (mental) representations “water” and “H_2O” have different *functional roles* because of their syntax (cf. Dretske 1981: 214-219). We care about that. So an analytic truth like

(3) Water is H_2O

although carrying no information, given Dretske`s explanation of information, is interesting since it connects two mental representations with a strong link which had not had that link before, if you did not know (3) before. A similar explanation applies to recognising consequences. These ideas on logical truths commit themselves to the representationalist/computationalist theory of the mind and await further elaboration.

4.2 An Ontological Solution

Even given *intensional isomorphy* in a semantic approach incorporating syntactic features, there are logical truths getting the same meaning although being distinct:

(4) $(\forall x)\ Raven(x) \supset (\exists x)\ Raven(x)$; and

(5) $(\forall x)\ Dog(x) \supset (\exists x)\ Dog(x)$

would be an example. According to the first syntactic approach mentioned, (4) and (5) would have the same meaning. That could be acceptable, since what you learn in terms of logic from (4) you can learn from (5) as well. If you want to make a distinction between even these sentences you need more than logical form. To solve these cases an *ontological* solution might be needed which refers to the constituents (resp. the referents of the constituents). Such an ontological solution would incorporate a more finely grained carving up of sentences or their referents. If you do not care about ontological plenty, you can distinguish (4) and (5) since the one contains the *property* of *being a dog* while the other contains the property *being a raven*. Situation semantics (Barwise/Perry 1983; Devlin 1991) is such a finely grained approach. For example the infon <<dog, fido, 1>> (Fido is a dog) and the infon <<dog, hasso, 1>> (Hasso is a dog) are different infons, since the first involves the *object* Fido while the latter involves the *object* Hasso. An analysis of compound infons and a consequence relation can then establish the difference between (4) and (5). This kind of solution would involve heavy ontological commitment.

§5 Hintikka´s Approach

Consider now Hintikka´s approach (Hintikka 1970, 1973). He was one of the first to address the problem as a problem of the information content of logical truths. He considered the problem in the light of epistemic modal logic, asking for the correctness of two principles of epistemic modal closure. Epistemic modal logics (i.e. epistemic logics of the early kind, modelled after alethic modal logic) should be normal modal logics if there should be any logic of the epistemic operators at all.

Normal modal logics contain a rule of *necessitation:*

(1) $\vdash \alpha \rightarrow \vdash \Box\alpha$

and the *K-Axiom:*

(2) $\vdash \Box(\alpha \supset \beta) \supset (\Box\alpha \supset \Box\beta)$

i.e. the *derived rule*: $\vdash(\alpha \supset \beta) \rightarrow \vdash(\Box\alpha \supset \Box\beta)$. The counterparts in epistemic modal logic are then:

(3) $\vdash \alpha \rightarrow \vdash K\alpha$ [all logical truth are known]

(4) $\vdash (\alpha \supset \beta) \rightarrow \vdash (K\alpha \supset K\beta)$ [all consequences of known premises are known]

which are considered highly contra-intuitive.

Hintikka tries to avoid these contra-intuitive consequences by distinguishing kinds of information: *surface* vs. *depth* information. But he also restricts the closure principles.

Hintikka believes that there is a sense of information in which logical inference can *add* to our information, i.e. our knowledge. His explication relates our problems in recognising a logical truth (i.e. in getting additional information) to the increasing depth of a procedure of checking quantificational consistency (in First Order Logic). *Surface* and *depth* information are defined relative to a nesting of quantifiers. Closure (under **K**) does hold only if $\alpha \supset \beta$ is a surface tautology at the depth of α (i.e. at the depth of what is already known). That is, we look at the depth of quantification in α and the depth of quantification in β; if the depth of β is not greater than that of α, Hintikka sees no problem and closure under **K** should apply. If the depth of β is greater than that of α, closure under **K** cannot be applied automatically. When we learn $\alpha \supset \beta$, we gain information (vis. the difference between surface and depth information). Increasing the depth and then detaching (in a conditional) can add to our knowledge. But closure (under **K**) does not apply here. An account of epistemic closure, therefore, depends on an account of logical depth information (in a first order possible worlds semantics).

Although it is difficult to explain Hintikka`s approach in detail, let us look at some of its features. We need some *measure* of surface and depth information to compare them. The *degree* of a formula is obtained as sum of the number of free singular terms and the maximal number of quantifiers whose scopes have a common part in the formula (i.e. its *depth*). Quantifiers are pushed inwards. *Depth* depends on quantifier changes like “$\exists x \forall y \exists w \forall v(\ldots)$“ (depth 4, say), since “$\exists x \exists y$” could be simplified into a single quantifier (on a pair). Checking for consistency is done depth by depth, looking for trivial inconsistency at the subordinate clauses´ depth (the subordinate clauses being one within the scope of another quantifier) by instantiating the variables bound by “$\exists$“. Like constituents *logical truths* get assigned a corresponding *depth* in the procedure. If you formulate these logical truths as conditionals you see which of them are information increasing.

This procedure is, of course (since First Order Logic is not decidable), not effective when applied to the non-finite case – which makes so checking the applicability of closure under **K** non-effective. *Given* that we know that α is a logical truth, counting its quantificational depth is effective. So determining the logical truth of a formula should be distinguished from determining whether it has an information increasing structure.

What can we say about Hintikka`s approach then? There seem to be quite a few open questions. Is this a *psychological* theory? Where from? It seems nobody employs these procedures or the corresponding measures. So let us assume it is a model for some unspecified process going on in assessing and recognising logical truths. The model may explain why information is gained by consequence, but it does not say *which* information we get if it were to be expressed in words. Why are just *quantifiers* the problem? Even though PC is decidable *we* might not be able to discover that some α is a tautology. So even closure within propositional epistemic logic is a problem. Why not simply say we do not know all the consequences of our beliefs, since this surpasses our capacities because of computational complexity (we have not enough time and storage) or – in some cases – undecidability? Although Hintikka employs the machinery of the semantic approach, the procedure looks cumbersome and non-effective. That might invite one to look for another approach (within situation semantics or some version of a syntactic approach).

§6 Algorithmic Information Theory to the Rescue?

Algorithmic Information Theory (Chaitin 1997a, 1997b) is a theory of information content, not of information flow. It deals with word strings. The basic measure is still bits, but Algorithmic Information Theory focuses not simply on the coding scheme but on matters of generating a word string by a program. A string has some measure in bits. The information content of a string is the *length of the shortest program (in bits)* which is needed to *generate* the string. The length of the shortest program for a string is also its *complexity*. A finite string of length n can be "programmed" by having it simply printed, with length n+k, k being the length in bits of the minimal code to print it. (The real problem are *infinite* strings, but since there are no infinite sentences this is no problem here.) A string is *random* if the size of the shortest program for it, if there is any, is not shorter than the string itself. *Most* strings are random, since there are more strings than well-formed programs. So the *great majority* of strings of length n are of complexity very close to n. So the basic definitions of interest here are:

(1) The complexity $I_C(s)$ of a binary string is defined to be the length of the shortest program p that makes the computer C output s, i.e. $I_C(s) = \min [lg(p)| C(p)=s]$

(2) A random binary string s is one having the property that $I(s) \approx lg(s)$.

The complexity $I_C(s)$ mentioned in (1) defines also the information content of a string. If you know its complexity you know the amount of information present in it.

Algorithmic Information Theory could be a *syntactic* solution at least to the problem why different logical truths have different information content. Logical truth – at least those which are theorems within a logical system – are *not random,* as one would expect, since by their very definition there are programs for them: one could assume that *one* program capable to generate a string that is a logical truth is it`s *proof*. So logical truth would have definite information content, and different logical truth could have different ones. (Given that we single out *that* program.) And given that we have found the shortest proofs of them we have the length of the proof available, so that we can see whether much or not so much information is gained in a logic course.

§7 Conclusion

Algorithmic Information Theory is another attempt for a syntactic solution. Its only shortcoming is that even though each logical truths *has* a proof, given the completeness of the system, we have not always a constructive procedure to *deliver* this proof, even if we know that the formula is a logical truths. I, therefore, prefer the approach hinted at in §4.1. Given some syntactic categories sentences are analysed in, it is a mechanically solvable problem to assign each sentence its most finely grained syntactical form. This applies also to logical truths and distinguishes them from each other. The presupposition of this approach, however, was that we not only have a general (truth conditional) account of meaning, but can also integrate a notion like *intensional isomorphy* within it. That is – to say the least – quite controversial, and cannot be argued for here. Even if this is to heavy a burden to shoulder, given some syntactic approach or even retreating to an ontological solution we have to save the phenomenon. We learn something by doing logic.

References

Barwise, Jon/Perry, John (1983). *Situations and Attitudes*. Cambridge/MA.

Bremer, Manuel (1993). *Epistemische und logische Aspekte des semantischen Regelfolgens.* Aachen.

Carnap, Rudolf (1955). "Meaning and Synonymity in Natural Languages", *Philosophical Studies*, 7 , S.33-47.

Carnap, Rudolf/Bar-Hillel, Joshua (1952). An Outline of a Theory of Semantic Information. Technical Report 247 Research Laboratory of Electronics MIT.

Chaitin, Gregory (1997a). *Algorithmic Information Theory*. New York, 3rd Edition.

- (1997b). *Information, Randomness and Incompleteness*. New York, 2nd Edition.

Devlin, Keith (1991). *Logic and Information*. Cambridge/MA.

Dretske, Fred (1981). *Knowledge and the Flow of Information*. Oxford, 2nd Edition.

Floridi, Luciano (forthcoming). “Outline of a Theory of Strongly Semantic Information”, forthcoming in *Minds and Machines*.

Fodor, Jerry (1987). *Psychosemantics*. Cambridge/MA.

- (1994). *The Elm and the Expert*. Mentalese and its Semantics. Cambridge/MA.

Hintikka, Jaakko (1970). “Surface Information and Depth Information”, in: Hintikka, J. /Suppes, P. (eds.) *Information and Inference*. Dordrecht.

- (1973). *Logic, Language-Games and Information*. Oxford.

Shannon, Claude/Weaver, Warren (1949). *The Mathematical Theory of Communication.* Illinois.