Natural Implication

by Sven Nilsen, 2019

In logic, material implication might produce non-intuitive results: When the assumption is false, anything follows. In this paper I show that when we say "A implies B" in natural language, we mean that "A", by how it is used in argumentation, implies that it is sound or true, but not necessarily both.

The problem starts when we say that `A` means something contradictive as an argument:

meaning_of("A") = (A => B) ∧ (A => ¬B)

For example:

- It looks like the Sun is orbiting the Earth (`A => B`).
- It looks like the Earth is rotating around its own axis (`A => ¬B`).

Naturally, we think of both statements as true. Yet, Logic states that if that is the case, then anything, including false itself, is true.

We think of "It looks like" as an argument "A", which can be used to explain why some people believe the Sun is orbiting the Earth or why some people are believing the Earth is rotating around its own axis.

Using the notation from Naive Zen Logic:

(B ? some people) => ((it looks like => B) ? some people)
(¬B ? some people) => ((it looks like => ¬B) ? some people)

Whenever something is believed by some people, it must be because it looked that way at first:

∀ X { (X ? some people) => ((it looks like => X) ? some people) }

This holds for people that are honest to themselves and intelligent.

If I was zen-consistent, then if I believe a smarter version of myself believes `X`:

∀ X { ((X ? .me) ? me) => ((it looks like => X) ? me) }

Then, I believe that if I observe something like `X`, then I believe `X`. Naive zen-consistency makes me just believe it without further evidence, but the above interpretation is also a non-naive zen-consistency in itself. I could swap zen-consistency with this rule and it could be used as an intermediate step toward zen-rational behavior, where smarter versions of myself might be a little wrong sometimes.

Zen-rationality often has to deal with problems that runs into logical impossible situations, but must be solved within a decent amount of time for practical problems. One way to start is to take natural language seriously on face value, without trying to impose a "correct" theory of semantics. So, with this motivation, our "It looks like" is an explanation for whatever is believed by some people.

The problem is in Logic interpreted as each proposition representing truth:

$$((A => B) \wedge (A => \neg B)) = (A = false)$$

What we mean by “It looks like” has a different interpretation than truth. If we try to substitute “A” with “It looks like”, we must also ask how we choose to think about `”It looks like” = false`. This is not what we meant by “A”, so either we limit ourselves of how we think, or we figure out another way.

The way we resolve this problem is by substituting “true/false” with “sound/unsound”:

$$((A => B) \wedge (A => \neg B)) = (A = unsound)$$

Since we substitute all truth values, the Logic is given a different interpretation than truth.

Soundness is isomorphic to Logic and therefore occupies the whole space of possible logical constructions. This means one can not mix truth and soundness within the same Logic.
However, one can create a combinatorial Logic that combines truth and soundness. While not being isomorphic to Logic, it is possible to encode it into Logic, which makes theorem proving possible.

A sound argument can be true, but it can also can be false (e.g. something physically impossible).
A true argument can be sound, but it can also be unsound (e.g. a fact that is irrelevant).

This constraint is what underlies the semantics of natural implication.

	A is true	**A is false**
A is sound	coherent	anti-coherent
A is unsound	anti-coherent	coherent

When the two logics of truth and soundness are the same, they are coherent:

$$coherent => \forall P \{ ((P = true) => (P = sound)) \wedge ((P = false) => (P = unsound)) \}$$

When the two logics of truth of soundness are opposites, they are anti-coherent:

$$anti\text{-}coherent => \forall P \{ ((P = true) => (P = unsound)) \wedge ((P = false) => (P = sound)) \}$$

Other states corresponds to a constraint of one or more propositions in one of the logics:

	A is true	**A is false**
A is sound	true-decoherent	false-decoherent
A is unsound	true-decoherent	false-decoherent

	A is true	**A is false**
A is sound	sound-decoherent	sound-decoherent
A is unsound	unsound-decoherent	unsound-decoherent

Natural implication imposes a such constraint as a relationship between two propositions of truth:

$$
\begin{aligned}
&\text{natural\_implication}(A, B) = \\
&\qquad ((A = \text{false}) \Rightarrow (((A \Rightarrow B) = \text{false}) \land ((A \Rightarrow \neg B) = \text{false}))) \lor \\
&\qquad ((A = \text{unsound}) \Rightarrow (((A \Rightarrow B) = \text{unsound}) \land ((A \Rightarrow \neg B) = \text{unsound})))
\end{aligned}
$$

This is logically equivalent to:

$$\text{natural\_implication}(A, B) = (A = \text{true}) \lor (A = \text{sound})$$

Now, since we proved that `( (A => B) ∧ (A => ¬B) ) = ( A = unsound )`, then "A" must be true:

$$\frac{(A = \text{unsound}) \land ((A = \text{true}) \lor (A = \text{sound}))}{A = \text{true}}$$

This fixes the intuition that whatever is believed by some people, the cause of it must be that it looks like the world is that way in the first place. We do not have to limit ourselves to e.g. use propositions only for truths. It is possible to construct a combinatorial Logic and do theorem proving just fine.

The issue here is what we mean when using words. "It looks like" is an arbitrary expression in natural language which may not correspond to a clean and coherent way of modeling the world. However, whatever we say in natural language should be formalizable. We want to study what we mean when we say things in natural language, as a theory of its own right, instead of just studying limited worlds such as a single Logic of truth.

Natural language has a more messy semantics that results from being designed to arrive at better world models over time. Logic forces you to be more precise upfront, but if you make a false assumption, you can derive anything. This would be widely unsafe in natural language, so by using a combinatorial Logic, it can check for errors.

In natural language, one can say that:

- It looks like the Sun is orbiting the Earth (`A => B`), and …
- It looks like Earth is rotating around its own axis (`A => ¬B`)

This is written in Logic:

$$(A \Rightarrow B) \land (A \Rightarrow \neg B)$$

If we take this argument to be true, then it means that "A" must be sound but false.
If we take this argument to be sound, then it means that "A" must be unsound but true.
This property means actually the same thing: Anti-coherence.

$$\text{anti\_coherent} := (\text{sound} \land \text{false}) \veebar (\text{unsound} \land \text{true})$$

The `⊻` operator means XOR. However, this can be simplified to:

$$\text{anti\_coherent} := \text{unsound}$$

Anti-coherence has the property that if natural implication is violated, then the argument is unsound:

$$\neg((A = \text{true}) \lor (A = \text{sound})) \quad \Rightarrow \quad A \Rightarrow \text{unsound}$$

Therefore, if `A` does not imply unsoundness, then natural implication is true:

$$\neg(A \Rightarrow \text{unsound}) \quad \Rightarrow \quad (A = \text{true}) \lor (A = \text{sound})$$

This means if you test for unsoundness in the conclusion, you do not have to add constraints at all!

Once more, I need to point out that `(A => B) ∧ (A => ¬B)` means that `A` is false, which is not what we mean when we say “It looks like”. While this argument is sound when it does not look like the Sun is orbiting the Earth, nor does it look like Earth is rotating around its own axis, it is not very helpful. We want to speak about what we see. So, `( (A => B) ∧ (A => ¬B) ) = sound` means `A` is unsound.

If it would not have been for the observations that “It looks like” was true in both cases, the argument would not have been unsound!

Either interpretation is valid in natural language, but since we exist in a real world, we can distinguish the valid interpretation from the invalid one. Naturally, it feels wrong to think that “It looks like” is false, so we use the interpretation that it is true, but as a logical consequence this is an unsound argument to distinguish between cases. The constraint of natural implication enforces this relationship, such that we must use the rules according to the use of natural language.

Natural language means how it is used. This definition of “truth” is different from how Logic is used from within natural language. Logic used from within natural language defines “truth” to be how we think about truth. Therefore, it is impossible to think about natural language the way Logic is used from within it. The following example will demonstrate this mind blowing property.

In natural language, **one can also say that**:

- It looks like the Sun is orbiting the Earth, and …
- The Earth is rotating around its own axis, so it does not look like the Sun is orbiting the Earth

Somehow, this **makes sense**, but it does not have anything to do about truth. From the sentences above, we understand that we are saying something that seems true and we are saying about the same thing that it is false. We are talking about two different perspectives.

This problem is common in natural language, because there are hidden states of information.

However, no matter how you think it makes sense, in Logic this is completely false:

$$(A \Rightarrow B) \land \neg(A \Rightarrow B) \quad = \quad \text{false}$$

To be more precise:

- In one way, it looks like the Sun is orbiting the Earth, and …
- In another way, it does not look like the Sun is orbiting the Earth

These ways of looking at the world correspond to different semantics of soundness. An argument might be sound in one place but unsound in another place.

Remember that Logic used from within natural language defines “truth” to be how we think about truth. Natural language itself uses a different definition of truth, so when we try to use Logic, we find that things making sense in one way or another suddenly does not make sense in Logic.

If we encoded the hidden state, which correspond to different soundnesses, it would look like this:

$$\frac{(\ (A \Rightarrow B) = \text{one\_way}\ ) \wedge (\ \neg(A \Rightarrow B) = \text{other\_way}\ )}{\text{one\_way} \neg= \text{other\_way}}$$

The problem is that this can not be deduced in advance: Hidden states are discovered afterwards.

What we meant to say above without the hidden state, is something sound-decoherent.

Sound-decoherence has this surprising property:

You can not say what you mean by sound-decoherence in Logic!

Why? Because Logic used within natural language is about truth, and there is no more room left!

Every sound-decoherent statement `X` has truth `X ∧ ¬X`, which is always false.

In natural language, one can also express a coherent model of the world:

- Either it looks like the Sun is orbiting the Earth, or …
- The Earth is rotating around its own axis, so it look like the Sun is not orbiting the Earth

This can be written down in Logic:

$$(A \Rightarrow B) \veebar (A \Rightarrow \neg B)$$

The `⊻` means XOR. This model is both true and sound, so it is coherent.
The choice of model is how we talk about things in a consistent way, meaning different things.

Now, we will back off a bit and consider these different ways of expressions from a bird perspective.

It might seem weird to think like this using Logic, but remember the following: A single Logic permits **anything** to follow from a false assumption, which means you can say whatever you want about the argument.

In path semantics the identity of the argument requires that everything we can say about it must remain the same. At first sight, it seems that path semantics can only describe true things, but that is not what it says. If you can figure out a way around false assumptions such that you can say things about specific versions of them that does not hold for every false assumption, then that is still a valid concept in path semantics. Path semantics is intuitive, because it is closer to natural implication than Logic.

For every proposition `P`, the following holds:

$$(\ P = (\text{true} \lor \text{sound})\ ) = P$$

Or, simply:

$$(\text{true} \lor \text{sound}) = \text{true}$$

If we assert that `P = sound`, we can prove whether some part of it is unsound.

We can not assert that `P = true`, because this is just `P`. We can not prove `P` from `P = sound`.

Howeever, there is a cool trick. We can do the following:

$$\frac{P = \text{sound}}{(P = \text{true}) \lor (P = \text{sound}) \qquad = \qquad P}$$

This might not seem very useful, but look what happens when we imply exclusivity:

$$\frac{(P \Rightarrow Q) \land (P \Rightarrow \neg Q) \qquad = \ \text{sound}}{(P = \text{true}) \lor (P = \text{sound}) \qquad = \qquad P}$$

Simplified:

$$\frac{(P \Rightarrow Q) \land (P \Rightarrow \neg Q) \qquad = \ \text{sound}}{P \lor (P = \text{sound}) \qquad = \qquad P}$$

Since `P` is unsound, it must be true. Otherwise, it would not be sound that `P` implies contradiction.

If we assert that some assumption that implies contradiction is sound, then natural implication becomes the fact that we observe, but which is not sufficient to use as argument to make distinctions.

Hence, natural implication makes it possible to reason consistently about “truth” the way it is used in natural language. It extends the Logic of truth to a combinatorial Logic of truth and soundness.

In natural language, every proposition that we use to imply something must use natural implication. Or else, you end up with false observations.

For previous work in this area, see published notes of Ludwig Wittgeinstein.