Review of: "[Research Note] A note on Hempel’s paradox"

Michele Loi

Potential competing interests: No potential competing interests to declare.

This article presents a way of thinking about Hempel's paradox that is new to me. The author describes the confirmation relation in inductive inference as “non-monotonic”. By this, he means: if one adds more premises, the validity of the previous inferences can change. This is contrasted with the validity of deductive inference, which is non-monotonic. If all humans are mammals and Socrates is human, then Socrates is a mammal. The validity of this inference cannot change if, let us suppose, it turns out that Socrates died when he was three hundred years old. According to the author, this property should not be taken for granted in inductive inference. Intuitively, after I see two swans, which are white, concluding that all swans are white is a valid conclusion. This conclusion is no longer valid after an encounter with a black swan.

An interesting application is that one may consider this process in reverse temporal order, that is, ask what follows from excluding a premise. Imagine travelling back in time. After seeing two white swans and one black swan, I cannot confirm “all swans are white”. But if I were to “exclude” my last observation (as if moving backwards in time), I could then say that, now as I'm back in the past, when I have only seen white swans, and limited to all the observations made so far, really, at this point, I can *legitimately* confirm that all swans are white.

Notice that I'm using a temporal interpretation of the idea of non-monotonicity. But the idea is much more general than that. The general concept is one of “context”. The author describes non-monotonicity as the idea that a certain inference ("all swans are white") may be valid relative to one context, but not relative to the other. Travelling in time is just one possible interpretation of a change in contexts.

Now let us revisit Hempel's paradox from this perspective. I see a black raven. How can we fail to get my belief that all swans are white confirmed to a degree? As I understand, the author's idea is that there is a context which **only** contains the observation “raven (a) Λ black (a)”. And relative to this context, I can legitimately confirm “non-white (a) Λ non-swan (a)”, to a degree. Moreover, there is a context which **only** contains the observation “non-white (a) Λ non-swan (a)” and relative to this context, I can legitimately confirm “all swans are white”, to a degree. But we shouldn't take for granted that the validity of

I. “non-white (a) Λ non-swan (a)” *confirms to a degree* “all swans are white”

in context A,

and the validity of

II. “raven (a) Λ black (a)” *confirms to a degree* “non-white (a) Λ non-swan (a)”

in context B

implies the validity of

III. “raven (a) Λ black (a)” confirms to a degree “raven (a) Λ black (a)” *that confirms to a degree* “non-white (a) Λ non-swan (a)”

in context C. (Notice that the fact that C is a different context follows from my earlier assumption that context A contains *only* the observation of a black raven; while context B contains only the realization that some thing is both white and not a swan; while context C contains *both* the observation of a black raven *and* the realization that some thing is both white and not a swan).

In other words, the non-validity of III is compatible with the validity of I. and II. taken separately. This is no more mysterious, the note seems to suggest, than the idea that the “there is a white swan”, “there is a white swan” (in context A) is able to confirm (to a degree) that all swans are white, while in context B, where one has already observed a black swan, the same observation may fail to confirm that all swans are black (even to a degree; that is to say, the observation of a black swan may undermine the capacity of the white swan observation to add any degree of validity to the claim that all swans are white).

If I read the note well, inductive inference where different claims are taken as valid across different contexts can be modelled conjunction by formal logic, i.e., the validity of P1 in context A and P2 in context B does not allow me to treat P1 Λ P2 as valid in context C. Hence, when I start with context-relative validity, I must be wary of my inability to generate new valid claims by adding together pieces of information that are valid in different contexts *in general*.

The problem with this approach, as I see it, is that the author remains silent about rules that are valid for carving out contexts. What makes context A a different context from context B? Unless one is given a rule to determine when two different observations belong to a context, and when they do not, the strategy counts as potentially fruitful, but its validity is hard to verify.

In other words, the proposed weakness of the proposal seems to be that it is too strong, in the sense that, unless circumscribed through some additional norm, it allows one to defend all sorts of absurdities. My issue with the proposal is a Popperian concern: with the appropriate choice of context, I may be able to defend all kinds of logical absurdities. And this is problematic insofar as it limits my ability to purge my beliefs from all sorts of absurdities, and the problem with losing that ability is that, as a result, nothing stands up, not just as valid in the sense of “scientific”; I can award myself infinite powers to defend my crazy body of beliefs from almost all forms of confrontation with an external world or other people.

Clearly, it may be ultimately false that I can distinguish contexts at will. So, with a strict enough law governing the admissibility of contexts, the objection I just advanced would not be valid. But the note does not give us such a strict enough law of context. Hence, in the absence of a law governing context, the validity of the idea seems hard to determine.

And additionally, I am not even sure that my charitably imagined strategy to avoid the “anything goes” conclusion - that is, having a rule for identifying contextual identity and difference - is compatible with the general idea. For what about the validity of the contextual norm? If the validity of the contextual norm is also defended inductively, then there must be an inductive relation of support that is entirely context-invariant. And how is this fact to fit with the general idea of non-monotonicity? On the other hand, if the validity of the contextual norm is *not* defended inductively, where does knowledge of the norm for identifying contexts derive from?

I can regard the approach envisaged as an interesting, and potentially instructive, one for dealing with this age-old dilemma; but I am unable to really say I'm convinced by it because I am not able to see (at this level of generality) how many absurdities one can derive from it. I remain agnostic about the validity of the proposal as a strategy to deal with Hempel's paradox until the author provides a rule governing the identification and differentiation of contexts.