UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

The Principle of Charity in Interpreting Scientific Theory: A Meta-Theoretical Polemic Against Theoretical Polemics

Permalink

Journal

ISSN

Author

Schroyens, Walter

Publication Date

2007

Peer reviewed

The Principle of Charity in Interpreting Scientific Theory:
A Meta-Theoretical Polemic Against Theoretical Polemics

Walter Schroyens (Walter.Schroyens@Ugent.be)

I proffer the 'principle of charity' as a fundamental guiding principle in interpreting theories communicated to us via scientific papers. The principle (named by Neil Wilson) conveys that when interpreting another speaker we should make those assumptions about his intelligence, knowledge, sense of relevance and so on, that will make most of what he says come out true. In Donald Davidson's work it often appears in terms of the injunction to optimise agreement between ourselves and those we interpret, that is, it counsels us to interpret speakers as holding true beliefs (true by our lights at least) wherever it is plausible to do. As David Vessey stated it; "the intuition behind the principle of charity is that if we interpret a sentence in such a way as to render it unintelligible, this is likely a sign of a poor interpretation rather than an accurate interpretation of an unintelligible view." Too often critics shun the principle of charity; thus often creating spurious contrast. Given my expertise, I here present an analysis of uncharitable interpretations made by critics of the mental-model theory; in part within the context of proffering an opposing theory.

"Basic conditionals have mental models representing the possibilities in which their antecedents are satisfied, but only implicit model for possibilities in which their antecedents are not satisfied" (Johnson-Laird & Byrne, p. 2002, p. 655). This *"implicit-model principle"* (a.k.a., initial-model principle) means that <if A then C> is *initially* represented by an explicit model of both <A> and <C> being true. Alternative possibilities <…> remain implicit: **{<AC>; <…>}.** Other processing principles articulate that when reasoners go beyond the initial representation, content and context aid the process of looking for alternatives to the initial model. Depending on context and context, people will reason *towards* different interpretations of the conditional. Johnson-laird and Byrne (2002, p. 650) also has the '*core meaning*' *principle*. "The antecedent of a basic conditional describes a possibility, at least in part, and the consequent can occur in this possibility. … The core meaning of *If A then C* is the *conditional* interpretation, which refers to the possibilities:"*{<A C>; <Not-A C>; <Not-A not-C>}.* This core-meaning principle in particular has received many uncharitable criticisms.

Any interpreter of Johnson-laird and Byrne (2002) must deal with the relationship between the implicit-model and core-meaning principle. The first one states that (*ab initio*) there is *one* explicit model; the latter states there are *three* explicit models. So what is it: one or three? Intellectual charity requires we come to an interpretation in which the principles are compatible. Without this we relegate Johnson-Laird and Byrne to self-contradictory inconsistency. We cannot presume any thinker is fully internally consistent (even though we should strive for such), but this self-contradiction is just too obvious. It is indeed easy to make sense of the relationship between the initial-model and core-meaning principle. They cannot hold at the same time. This immediately suggests the solution; they cannot apply simultaneously so, hence (being charitable), they do not apply simultaneously. That is, people have to start with either three models or one model and accordingly add models (when starting with one model) or delete models (when starting with three models). The label 'initial model principle', already suggests people start with the implicit-model representation: "The mental-model theory assumes that people reason from their understanding of a situation and that their *starting point is accordingly a set of models -- typically a single model of a single situation ...*".

Some have dealt with the one-vs.-three model problem by ignoring the implicit-model principle (and thus creating a straw-man theory). These critics suppose people start with the three models of material implication, i.e. the core meaning interpretation of if: "an individual might believe strongly … the [not-A_not-C] possibility. This is consistent with the Johnson-Laird and Byrne (2002) core semantics and consistent with background knowledge, *so they have no reason to delete this possibility.*" It grounds mistakes such as: "According to the extensional semantics of Johnson-Laird and Byrne (2002), the conditionals 'if p then q' and 'if p then not q' are *not* incompatible". *Ex hypothesis* the implicit-model principle these conditionals *are* clearly incompatible. The respective model sets are: *{<p q>; <...>}* and *{<p not-q >; <...>}*. The two explicit models cannot be integrated with one another into a single model: Something cannot be possible and impossible at the same time. That is, people who satisfice will consider the two conditionals incompatible. Others might not forget the implicit model, and fill the empty placeholder <....> with the explicit model of the other conditional. Another uncharitable interpretation, grounding straw man arguments is: "under the material conditional, the probability of a conditional [if A then C] and its contrapositive [if not-A then not-C] must be the same". The initial-model sets are: *{<A C>; <…>}* and *{<not-A not-C>;<…>}*. It follows we cannot predict that the probability of these conditionals is the same. For those who satsifice with the initial model it will clearly not be; hence overall it cannot be the same either.

When a theoretical ambiguity is readily resolved, theorists should do exactly that (independent of who 'owns' the theory). It yields a situation in which one is analysing the strongest version of a theory and not some weak straw-man version that is easily falsified. Critiques of strong (charitably interpreted) theories are more likely to stick because many subsidiary assumptions, *ceteris paribus* clauses and other 'escape hatches' for a potential counter-argumentation have already been dealt with in interpreting the theory charitably. Charitable interpretations are thus more likely to advance our theoretical insights.

References

Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. Psychological Review, 109(4), 646-678.