The Incoherence of Heuristically Explaining Coherence

Iris van Rooij (i.v.rooij@tm.tue.nl)
Human-Technology Interaction, TU Eindhoven,
Den Dolech 2, 5612 AZ Eindhoven, The Netherlands

Cory D. Wright (cory@mechanism.ucsd.edu)
Depts. of Philosophy & Cognitive Science, UCSD
9500 Gilman Drive, La Jolla, CA 92093-0119 USA

Keywords: coherence, abduction, heuristics, constraint satisfaction, NP-hard, explanation, truth.

Advancement in cognitive science depends, in part, on doing some occasional "theoretical housekeeping." In this poster, we highlight some conceptual confusions lurking in an important attempt at explaining the human ability for rational or coherent thought: Thagard and Verbeurgt's computational-level model of humans' capacity for making reasonable and truth-conducive abductive inferences (1998; Thagard, 2000). T&V's model assumes that humans make such inferences by computing a coherence function (f_{coh}), which takes as input representation networks and their pair-wise constraints and gives as output a partition into accepted (A) and rejected (R) elements that maximizes the weight of satisfied constraints. We argue that their proposal gives rise to at least three difficult problems.

Being NP-Hard Results in a Dilemma

T&V proved that f_{coh} is NP-hard. This result proves that there does not exist any efficient (polynomial-time) algorithm for computing f_{coh}, under widely-held assumptions in mathematics (Garey & Johnson, 1979). Insofar as they take cognitive feasibility to require efficiency (1998, p. 7), T&V cannot maintain that human minds can compute f_{coh} for all logically possible inputs (van Rooij, 2003). Hence, a dilemma arises: either (i) one must conclude that f_{coh} does not adequately characterize how representation networks comply with maximum coherence, or (ii) one needs to explain what special property real-world inputs have, such that humans can efficiently compute f_{coh} for those inputs.

Heuristics are Incoherent Explanations

T&V reject (i), but fail to recognize that doing so commits them to (ii). Instead, they misestimate their goal as being the design of inexact procedures (heuristics) to serve as (approximate) explanations of how people compute f_{coh}. This approach rests on a mistake—one that confuses the goal of *explaining how a computation is achieved* with the goal of *attempting to achieve a computation*. With their heuristics approach, T&V seem to avail themselves of the latter goal; but the goal should instead be the former, given (ii). For if theorists intend to explain *how* a computation is achieved, then the procedure posited at Marr's algorithmic level had better be an exact algorithm for the function posited at Marr's computational level (Marr, 1982), since the two levels otherwise make for incompatible and competing forms of explanation. To see why, consider a heuristic H that computes a function f_H. Because H is an inexact algorithm for f_{coh}, there exist inputs i such that $f_{coh}(i) \neq f_H(i)$. But then, the hypothesis that f_{coh} adequately characterizes human inference, and the hypothesis that H adequately describes the process by which humans make abductive inferences, are incompatible hypotheses for all those inputs where $f_{coh}(i) \neq f_H(i)$. Consistency and coherence demand that one of the two hypotheses be rejected.

Coherence Allows Contradictions to be True

It is well-known that representation networks can be highly coherent and internally consistent without necessarily tracking how the world actually is. As such, coherence theories of truth and justification are beset by problems of circularity and being ungrounded. Contrary to what T&V claim (1998, p. 2), we argue that f_{coh} fails to overcome these problems. In particular, we show that the model does not preclude opposing and mutually exclusive belief systems to be equally and maximally coherent, because there can exist partitions (A, R) and (A', R') that each satisfy a maximum number of constraints, while $A = R'$ and $R = A'$ (i.e., all elements accepted as *true* in the first partition are rejected as *false* in the second, and vice-versa). Prima facie, this result implies the absurdity that, for any statement p, cognizers are as justified in believing that p is true as they are in believing that p is false. This absurdity infects the model as an account of warranted assertibility; and since circularity is not avoided by changing how constraints are processed (e.g., from sequential to parallel processing), it seems that invoking f_{coh} will be insufficient to ground the human capacity to achieving true, justified belief.

References

Garey, M. R. & Johnson, D. S. (1979). *Computers and intractability: A guide to the theory of NP-completeness.* New York: Freeman.

Marr, D. (1982). *Vision.* San Francisco: W.H. Freeman and Company.

Thagard, P. (2000). *Coherence in thought and action.* Cambridge, MA: MIT Press.

Thagard, P. & Verbeurgt, K. (1998). Coherence as constraint satisfaction. *Cognitive Science*, *22*, 1–24.

van Rooij, I. (2003). *Tractable cognition*. Ph.D. thesis, University of Victoria, Victoria, Canada.