The Convergence of Explanatory Coherence and the Story Model: A Case Study in Juror Decision

Michael D. Byrne
School of Psychology
Georgia Institute of Technology
Atlanta, Georgia 30332-0170
byrne@cc.gatech.edu

Abstract

This paper presents an integration of two approaches to complex decision-making from very different traditions: from the psychology of jury decision, the Story Model, and from the philosophy of science, the Theory of Explanatory Coherence and its computational instantiation, ECHO. The subjects in Pennington & Hastie (1993) generated causal "stories" to represent the events related to a particular trial. These stories were modeled with ECHO, and ECHO reached the same verdicts as did the human subjects. The ECHO simulations were also linked to the trial testimony, which, despite the inconsistent nature of the testimony, actually increased the coherence of stories for two jurors with very different verdicts. Implications for both the Story Model and ECHO are discussed.

Introduction

One of the questions confronting both psychology and philosophy is understanding how it is that people make decisions in complex situations. Such situations often contain contradictory evidence, gaps in what is known, and the like. While no complete account has yet been offered, similar frameworks for understanding complex decisions have arisen, and from very different traditions. One tradition is that of the historical philosophy of science, which attempts to understand how it is that scientists come to accept new paradigms. While there have been many approaches to understanding what Kuhn (1962) termed a "paradigm shift," the one considered here is Thagard's (1989, 1992a) Theory of Explanatory Coherence (TEC) and the associated computational model, ECHO. The second tradition is the psychology of jury decisions, which attempts to understand how it is that jurors arrive at a particular verdict. This, too, is not a field with a unitary way of understanding its phenomena, but one model that has been particularly promising in this area is the Story Model (Pennington & Hastie, 1986; 1993). Interestingly, work beginning in these seemingly disparate domains has converged to the point that it should be possible to integrate the two frameworks into a unitary theory of complex decision-making.

While presenting such an integrated theory is beyond the scope of this presentation, it is possible to demonstrate that these two approaches are both consistent and complementary. This will be made clear by presenting TEC-ECHO simulations of some of the data which has been used to support the Story Model. To understand these examples, the two frameworks will first be described and then the simulations presented. A discussion of the implications for both approaches and possible integration will follow the presentation of the simulation models.

Explanatory Coherence and ECHO

In attempting to understand how it is that scientists make decisions to accept new theoretical positions, Thagard has constructed the Theory of Explanatory Coherence (TEC) and a computational modeling system which embodies the principles of TEC, ECHO. TEC and ECHO have been used to provide "a mechanism that can lead people to abandon an old conceptual system and accept a new one" (Thagard, 1992a, p. 62).

In TEC, "explain" is taken to be a primitive relation between propositions (P, Q, and $P_1..P_n$) in an explanatory scheme (S). Coherence, then, is the extent to which the propositions in the system follow the "principles" of explanatory coherence:

Principle 1. Symmetry
(a) If P and Q cohere, then Q and P cohere.
(b) If P and Q incohere, then Q and P incohere.

Principle 2. Explanation
If $P_1...P_m$ explain Q, then
(a) For each P_i in $P_1...P_m$, P_i and Q cohere.
(b) For each P_i and P_j in $P_1...P_m$, P_i and P_j cohere.
(c) In (a) and (b) the degree of coherence is inversely proportional to the number of propositions $P_1...P_m$.

Principle 3. Analogy
If P_1 explains Q_1, P_2 explains Q_2, P_1 is analogous to P_2, and Q_1 is analogous to Q_2, then P_1 and P_2 cohere, and Q_1 and Q_2 cohere.

Principle 4. Data Priority
Propositions that describe the results of observations have a degree of acceptability on their own.

Principle 5. Contradiction
If P contradicts Q, then P and Q incohere.

Principle 6. Competition
If P and Q both explain a proposition P_i, and if P and Q are not explanatorily connected, then P and Q incohere. Here P and Q are explanatorily connected if any of the following conditions holds:
(a) P is part of the explanation of Q.
(b) Q is part of the explanation of P.
(c) P and Q are together part of the explanation of some proposition, P_j.

Principle 7. Acceptability
(a) The acceptability of a proposition P in a system S depends

on its coherence with the propositions in S.
(b) If many results of relevant experimental observations are unexplained, then the acceptability of a proposition P that ex - plains only a few of them is reduced.

which were taken directly from Thagard (1992a). The coherence of a large system of explaining and contradicting propositions cannot be computed simply by informally applying the principles of TEC (despite criticisms to the contrary, which have not demonstrated themselves to be successful). In light of this, a connectionist (though not PDP) system called ECHO has been developed which makes this computation straightforward. TEC and ECHO have been used to explain numerous scientific revolutions such as the Copernican revolution (Nowak & Thagard, 1992; Thagard, 1992a) as well as various complex decisions, such as Hitler's belief the Allies would invade Calais rather than Normandy and the decision of the captain of the *USS Vincennes* which led to the destruction of a passenger aircraft (Thagard, 1992b). Most closely related to the present issue, ECHO has been used to model prominent jury verdicts (Thagard, 1989), though this work did not model the decisions of individual jurors, nor were the ECHO models based on explanations actually provided by human subjects.

One of the criticisms that has been raised (e.g. Giere, 1993) about the ECHO simulations is that the "explanations" have been provided to the system by the programmer—that is, the propositions used and the explanatory and contradictory links between them have all been decided upon by the same person, and since it is impossible to know what explanations that, say, Darwin actually considered, the simulations are in some way invalid.[1] The simulations presented here later address this issue by using the explanations provided by the jurors themselves rather than explanations provided by the programmer.[2]

The Story Model

The Story Model was developed to explain how individual jurors reach particular verdicts. Decision-making by jurors represents an interesting and complex psychological domain, because jurors typically receive large amounts of often contradictory evidence in essentially random order. One of the reasons for the success of the Story Model is that other, more "traditional" decision models have difficulty modeling human decisions under such conditions (Pennington & Hastie, 1992). The Story Model maintains that jurors arrive at decisions as the result of a three-stage process:

(1) Story Construction, "an active, constructive comprehension process in which they make sense of trial information by attempting to organize it into a coherent mental representation" (Pennington & Hastie, 1992, p. 190). These representations typically take the form of stories with causal links between episodes in the story. It is possible for jurors to construct more than one story, and in that case stories are judged on the basis of their acceptability. According to the Story Model, acceptability is a function of coherence, completeness, and uniqueness. These principles, when explained in greater detail, parallel those of TEC, as has been observed both by Thagard (1989) and Pennington and Hastie (1993).

(2) Verdict Representation, in which the juror constructs a representation of the possible verdicts. In most criminal cases, verdicts consist of more than simply "guilty" or "not guilty." For example, in the murder case used in Pennington and Hastie (1986), the jurors have four options: first-degree murder, second-degree murder, manslaughter, and not guilty. Verdicts are represented along four axes: *identity* (i.e. was the defendant the one?), *mental state* of the defendant at the time, *circumstances* during the event, and the *actions* taken by the defendant. While jurors differ from one another in terms of their representations of the verdicts, this does not play a central role, as differences in verdict representations are not associated with differences in decision outcomes (Pennington & Hastie, 1986).

(3) Story Classification, in which the story constructed in step 1 is matched to the verdicts represented in step 2. The central element here is the goodness of fit between the story and the various verdicts. The verdict with the best fit to the story is hypothesized to be the one chosen by the juror.

The relationship between the Story Model and ECHO is clear: ECHO provides a computational account of the acceptability of the stories constructed and the story classification processes, and, to the extent that juror's decisions are in accord with ECHO predictions, supports the psychological plausibility of ECHO. To demonstrate this more conclusively, I constructed ECHO simulation models based on the causal stories of two of Pennington and Hastie's (1993) subjects. These ECHO models do indeed reach the verdicts that the subjects reached, and have other interesting properties.

Simulation Models

The Jurors

The jurors simulated are Jurors 109 and 128, taken from Pennington and Hastie (1993). The jurors were presented with videotaped testimony, and then presented with four alternatives: first-degree murder, second-degree murder, manslaughter, and not guilty. The fictional case involves Frank Johnson killing Alan Caldwell. Since Johnson admitted that he killed Caldwell, there is no doubt about the identity or basic actions of the defendant. However, the exact verdict is not well-constrained by the testimony. For example, it is not clear if Caldwell attacked Johnson with a razor immediately before Johnson stabbed Caldwell. Beliefs about events such as that can play a key role in juror decisions.

[1] This is not a particularly imaginative criticism, as the algorithms and input data used by almost all computational models are supplied by the researchers—for example, the "feature vectors" supplied to connectionist models (e.g. Churchland, 1989). The burden is upon the critics to demonstrate why the explanations supplied ECHO are wrong. So far, Thagard's critics have failed to do this for all but the simplest of the ECHO simulations, the Darwin example (Thagard, 1989; 1992a, Chapter 6).

[2] It should also be noted that Thagard (1992a, Chapter 4) directly addresses this criticism in several different ways.

Figures 1 and 2 are reproductions of the Pennington and Hastie's (1993) Figures 3 and 4 (pp. 144, 145), which represent the stories generated by Juror 109 and 128, respectively. In these figures,

> [e]vents and episodes are represented by solid circles and the diameters of the circles indicate the degree of elaboration provided of events by the jurors; broken circles represent the defendant's goals, inferred by the juror. The arrows connect events that were explicitly linked by causal relations in the juror's verbal report. The letters J and C refer to the defendant Johnson and the victim Caldwell respectively.

It is important to note that Juror 109 delivered a "not guilty" verdict and Juror 128 delivered a "first-degree murder" verdict, after both of them had seen exactly the same evidence presented in exactly the same manner. The clear difference between the two jurors was in the stories they generated to explain the events which led to the trial.

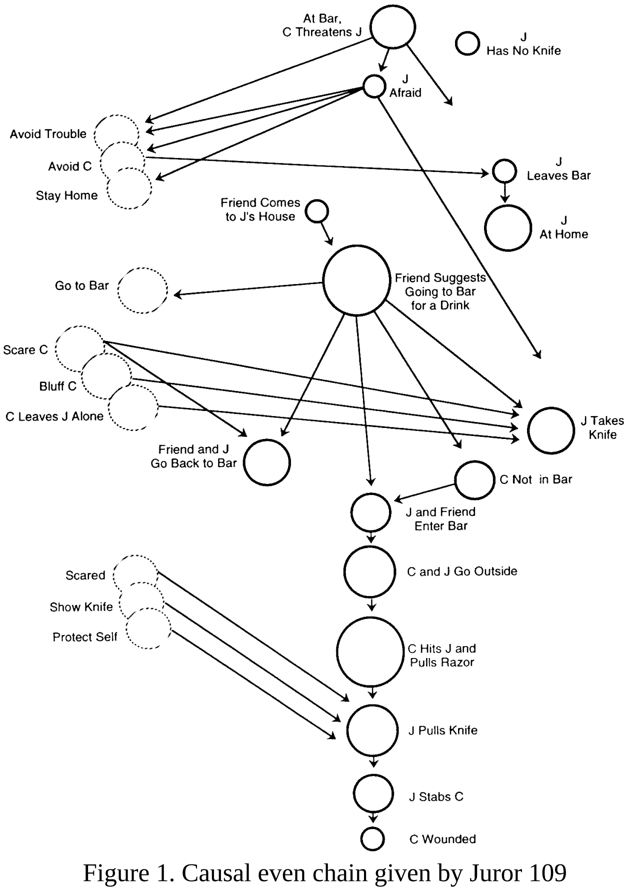

Figure 1. Causal even chain given by Juror 109

Simulation Elements

The ECHO simulation models constructed for both jurors contained many of the same elements, in particular those related to the verdict categories and the trial testimony. The verdict categories of action, mental states, and circumstances (taken from Pennington and Hastie, 1986) yielded 18 propositions and 7 explanations or contradictions. Linking the verdict categories and the final verdicts required four explanations and one contradiction.

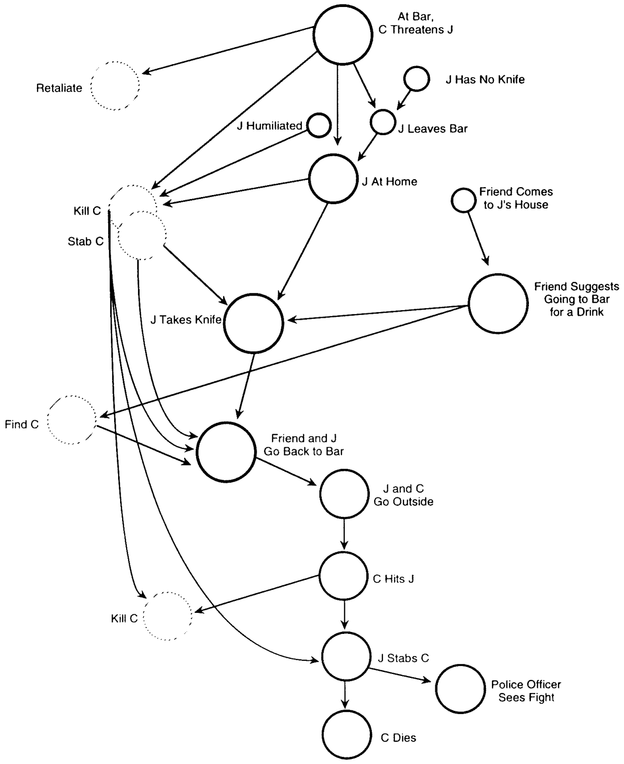

Figure 2. Causal event chain given by Juror 128

The four possible final verdicts generated four propositions. Since those verdicts are generally exclusive, five contradictory links were generated between pairs of final verdicts. The final verdict delivered by the juror was assumed to be the final verdict proposition with the highest activation at the end of the ECHO run, since ECHO activation level is intended to correspond to belief strength.

Nineteen propositions were generated by the testimony itself; while there were certainly more than 19 propositions in the testimony, those that seemed the most directly relevant were used. There were several pieces of directly contradictory testimony and these were included to see how ECHO would handle the contradictions.

Since one of the aims of this paper was to remove the supposed "programmer bias," and the stories of the two jurors included few references to the actual evidence, there were two models made for each juror: one including the testimony propositions and one without the testimony, as it is not guaranteed that the links made from the evidence to the story are exactly the ones made by the jurors. While this does have some impact on the ECHO network as a whole, the impact on the final decisions was negligible.

Juror 109's story consisted of 26 propositions and 18 explanations, all of which were again derived from the causal graph (Figure 1). Nodes from Figure 1 were represented in ECHO by propositions and links in the graph by ECHO explanations. Seven explanations/contradictions connected Juror 109's story to the verdict categories, and 19 more were necessary to connect the story to the testimony.

The story for Juror 128 consisted of 19 propositions and

14 explanations, all of which were derived directly from the graph presented in Figure 2 just as they were derived from Figure 1 for Juror 109. Eight explanations/contradictions connected Juror 128's story to the verdict category propositions. It should be noted that one of the propositions in Juror 128's story actually was a piece of the testimony, so only 18 additional explanations or contradictions had to be added to link the story to the 19 testimony propositions.

Summary of Results

A general summary of the results of the four simulations can be found in Table 1. This table presents the final (asymptotic) activation values for the total network and the four propositions representing the possible final verdicts. (Activation values range from +1 to -1 for a given proposition, with +1 indicating complete acceptance and -1 complete rejection.)

There are a several things to note about the simulation results. First and foremost, the simulations are in agreement with the verdicts reached by the jurors that were modeled. Second, according to ECHO, both of the stories constructed by the jurors are coherent explanations. This is important in that the explanations used were those constructed by the jurors themselves and not the ECHO programmer. Third, both explanations become even more coherent when related to the testimony. This is particularly interesting since the two stories yield opposing verdicts, and the testimony presented is not itself consistent. Both stories formed by the jurors integrate this contradictory testimony in a coherent way, even though the stories themselves differ dramatically.

Another interesting facet of the ECHO models relates to the testimony. Since the jurors did not observe any of the events as they happened, they must rely on the testimony and their own inferences to guide them. In many legal cases, though, testimony is somewhat less than guaranteed to be an accurate description of the events that took place. In the case examined by these jurors, the defendant and one of his best friends are also witnesses. Are they to be believed? As it turns out, whether or not the witnesses are believed depends on the content of their testimony. According to the ECHO simulations, testimony will be believed to the extent that it is coherent with the story that the juror constructs. In these simulations, for example, almost all of the defendant's testimony ends up with negative activation values (is not believed) for Juror 128, and all of it ends up with positive activation values for Juror 109. This is consistent with many of the ECHO simulations of scientist's beliefs, wherein certain experiments are considered "anomalies" and not believed by the scientists.

Table 1. ECHO simulation results

	Juror 109		Juror 128	
	story only	with testimony	story only	with testimony
Total coherence	.37	.59	.49	.78
Not guilty by self-defense	.34	.40	-.54	-.54
Manslaughter	-.47	-.40	-.26	-.28
Second-degree murder	.10	-.09	.49	.49
First-degree murder	.12	.03	.56	.57

Discussion

Despite their different origins, ECHO and the Story Model can work together to provide a compelling account of how people make complex decisions. While this account would certainly be more compelling with more jurors, in particular those delivering manslaughter and second-degree murder verdicts, the results presented here are promising. This has implications for further work on both the Story Model and ECHO.

Implications for the Story Model

One of the primary advantages for the Story Model of the ECHO approach is that it is less post-hoc than the present Story Model. As it stands, the Story Model is more an account than a predictive model (but see Pennington & Hastie, 1992). Jurors' stories are noted to be consistent with the verdict categories after the final verdict from each juror is known. Because "coherence, completeness, and uniqueness" are not formally defined in the Story Model, multiple interpretations of any given story are possible. With ECHO, the coherence is computed for a single verdict, making the prediction clear.

Second, the Story Model has been applied primarily to the domain of juror decisions. While, in principle, the Story Model is part of a more general framework of explanation-based decision making, most of the work on explanation-based decision making has been conducted as work on the Story Model. While this is certainly reasonable given the complexity of the task confronting jurors, the success of ECHO in domains outside of juror decision bodes well for the extension of the Story Model to other domains.

Implications for TEC-ECHO

One of TEC-ECHO's more caustic critics is Glymour (1992), with two major points: ECHO lacks psychological plausibility and the complex algorithm used by ECHO is unnecessary.[3] This fusion with the Story Model addresses both of these criticisms. Glymour (1992, p. 470) claims that "there is no psychological case at all" for the way ECHO computes coherence. While this claim ignores other successful applications of ECHO to psychological data (Ranney & Thagard, 1988; Schank & Ranney , 1991), this criticism is rendered even weaker by the present work. Jurors do indeed appear to make decisions that are consistent with the ECHO simulations, giving further support for ECHO's psychological plausibility.

The support for ECHO would be stronger if the stories that individual jurors rejected were also included and shown to

[3] Thagard (1992c) addresses this latter criticism quite effectively, this work merely serves to provide further evidence in favor of ECHO's algorithm.

have lower coherence than the story each juror decided upon. Another point in support would be if two jurors with contradictory stories were brought together and the juror with the story having greater total coherence “won” out (Juror 128 in this case). In fact, such an enterprise would be quite useful, extending both the Story Model and ECHO to the domain of complex decision-making by groups and not just individuals.

Glymour’s second criticism is addressed by this work as well. Glymour’s “pocket calculator” algorithm (1992, p. 474) for ECHO has a critically linear aspect to it which is not found in ECHO. While it may indeed agree with ECHO that the jurors’ stories are coherent and yield the decisions they do, it is unclear that Glymour’s algorithm will yield increases in coherence for both stories given the inconsistent nature of the testimony. Again, until Glymour can demonstrate a simpler algorithm that yields the consistency of results that ECHO does, there is no reason to believe that Glymour’s criticism is a valid one.

What Is an Explanation? One of the criticism that has been leveled at TEC-ECHO by both the previously-mentioned critics (Giere, 1993; Glymour, 1992) is that ECHO begs the question of what an explanation is. When “P explains Q” is provided in the context of TEC, what does “explain” actually mean? Are all explanations the same? Thagard (1992a) attempts to address this question with the answer that explanations take a variety of forms. Explanation, Thagard maintains, is a complex process that can include suprocesses based on deductive, statistical, schematic, analogical, causal, or linguistic/pragmatic subprocesses. There is no single way to construct an explanation, and the “goodness” of an explanation is a function of the explanatory system in which it is embedded.

This is entirely consistent with the data provided in Pennington and Hastie (1993). The inferences which connect one part of their story with the next take a variety of forms, all of which are equally valid for that juror. In fact, several of the explanatory links shown in Figures 1 and 2 are broken down by Pennington and Hastie (1993) to more primitive inferences, each of which could also be analyzed with ECHO (e.g. Pennington & Hastie’s (1993) Figures 5 and 6). Thus, there is no single answer to what an explanation is across all individuals, but once the (local) explanations have been formed, a given system of explanations seems to match the predictions made by TEC-ECHO. While this may be something of a difficulty for ECHO as a normative model, it provides healthy support for ECHO as a predictive one.

Conclusions

Understanding how people make complex decisions is a critical question for both psychology and philosophy, and an approach which integrates detailed analyses of the complex explanations formed and the coherence of those explanations could potentially shed a great deal of light on the problem. There is still plenty of work to be done here, of course, particularly in the area of understanding exactly how people construct these causal stories, but the integrated Story Model/ECHO approach offers much promise in answering the question of how people make decisions in complex situations.

Acknowledgments

I would like to thank Paul Thagard for providing me with the Macintosh implementation of ECHO, Dave Rettinger for providing me with reprints of most of the Story Model work, and Nancy Nersessian for getting me started on this project. I would also like to gratefully acknowledge the financial support of the National Science foundation through its Graduate Fellowship program.

References

Churchland, P. M. (1989). *A neurocomputational perspective: The nature of mind and the structure of science*. Cambridge, MA: MIT Press.

Giere, R. N. (1993). Explaining conceptual revolutions. Unpublished manuscript, presented at the Eastern Division meeting of the American Philosophical Association, Dec. 28–30, 1993.

Glymour, C. (1992). Invasion of the mind snatchers. In R. N. Giere (Ed.) *Cognitive models of science* (pp. 465–474). Minneapolis, MN: University of Minnesota Press.

Kuhn, T., (1962). *Structure of scientific revolutions.* Chicago: University of Chicago Press.

Nowak, G., & Thagard, P. (1992). Copernicus, Ptolemy, and explanatory coherence. In R. N. Giere (Ed.) *Cognitive models of science* (pp. 274–309). Minneapolis, MN: University of Minnesota Press.

Pennington, N., & Hastie, R. (1986). Evidence evaluation in complex decision making. *Journal of Personality and Social Psychology, 51,* 242–258.

Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the story model for juror decision making. *Journal of Personality and Social Psychology, 62,* 189–206.

Pennington, N., & Hastie, R. (1993). Reasoning in explanation-based decision making. *Cognition, 49,* 123–163.

Ranney, M., & Thagard, P. (1988). Explanatory coherence and belief revision in naive physics. In *Proceedings of the Tenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum.

Schank, P., & Ranney, M. (1991). Modeling an experimental study of explanatory coherence. In *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum.

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences, 12*, 435–502.

Thagard, P. (1992a). *Conceptual revolutions*. Princeton, NJ: Princeton University Press.

Thagard, P. (1992b). Adversarial problem solving: Modeling an opponent using explanatory coherence. *Cognitive Science, 16,* 123–149.

Thagard, P. (1992c). Computing coherence. In R. N. Giere (Ed.) *Cognitive models of science* (pp. 485–488). Minneapolis, MN: University of Minnesota Press.