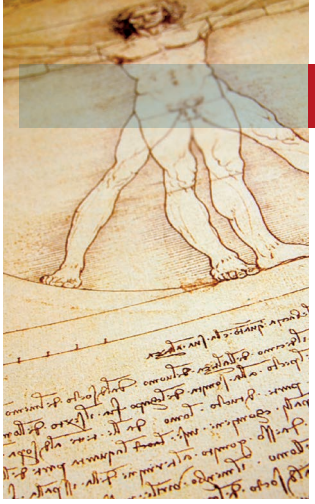

Editor: **Robert R. Hoffman,** Institute for Human and Machine Cognition, rhoffman@ihmc.us

Explaining Explanation, Part 2: Empirical Foundations

Robert R. Hoffman, *Institute for Human and Machine Cognition*
Shane T. Mueller, *Michigan Technological University*
Gary Klein, *MacroCognition, LLC*

Much of the literature in cognitive psychology and in the psychology of science focuses on causal reasoning on the part of scientists, especially about physical causation.[1,2] Scientists usually undertake investigations into determinate problems where there is a chance of making a discovery. But most people typically engage in causal reasoning about indeterminates. Why did the American military situation in Iraq improve from 2004 to 2008? Why did a certain sports team beat another in a championship game? There are no single or uniquely correct answers to such questions. Researchers such as James Reason have shown that accidents do not have single causes, so the quest for some single "root" cause or a culminating cause is bound to be an oversimplification and a distortion.[3]

In many contexts of decision making, the causes are often multiple, vague, and indeterminate. Frequently people never figure out actual or final causes. People sometimes stop their investigations at a fairly shallow level, demonstrating the reductive tendency.[4] Sometimes, reasoning about effects leads to a realization that the effects one is trying to explain have morphed. Time lags between cause and effect are inevitable; they create an additional layer of complication,[5] not simply because of the time but because intervening events cloud the picture. People still have to engage in causal reasoning under these conditions, but their reasoning will not follow the models of philosophers and scientists that lead to some single, final point where causal reasoning stops because the cause has been determined and the explanation of events is complete.

Reaching back to David Hume's analysis of causation, there appear to be three primary criteria that potential causes need to satisfy: propensity, reversibility, and covariation.[6] Propensity refers to the plausibility that a cause may have actually resulted in the effect. For example, if I start smoking today and at my annual physical exam tomorrow I am diagnosed with lung cancer, it is not plausible that my smoking led to the lung cancer. Reversibility, often referred to as mutability, is that the effect should disappear if the putative cause disappears. In modern terminology, this is counterfactual reasoning.[7] Covariation is the observed coincidence of causes and effects. For example, just over a century ago health experts tried to control yellow fever in Havana by reducing the mosquito population in part because of the correlation between the two, even in the absence of a plausible causal story. Manipulability has been suggested as a fourth criterion—if we manipulate a potential cause we should modify or alter the effect. The manipulability criterion runs into logical problems of circularity, but it might have value at the psychological level for describing how people assign causal attributions.

In applying such criteria, we need to take context into account. For example, if I use a hammer to shatter a watch crystal, we might conclude that the wielding of the hammer caused the crystal to shatter. But if the activity took place in a factory producing watch crystals, and the force of the hammer was carefully calibrated to test the individual crystals, then we might alter our causal account, speculating that perhaps the hammer was poorly calibrated or that the crystal was defective.

1541-1672/17/$33.00 © 2017 IEEE
Published by the IEEE Computer Society

As valuable as philosophical analysis might be, what forms of causal reasoning do we actually see when we adopt a naturalist method and look out at the world of human activity and explanation?

Purposes of Causal Reasoning

One of our experiments involved interviewing 10 specialists in logistics, intelligence, and command and control. They were asked to recount experiences in which it had been hard for them to figure out what was causing what (for example, why a fire truck did not receive its scheduled maintenance).[8] The most surprising finding involved situations where causal reasoning did not follow the course one would expect. There were situations where there weren't clear-cut beginnings or endings to the causal reasoning. The main issue for satisficing in causal reasoning is not determining a stopping rule (for example, when to stop the search for a causal explanation), but the condition under which the search for a causal explanation does not even start. We encountered situations that appeared ripe for causal investigation, only to find that the informant never bothered, usually because neither the job nor closure on the immediate task required a causal investigation.

Therefore, one of our first questions was, why do people seek causal explanations of events in the first place? It is often assumed in the literatures on explanation and causal reasoning that the purpose of coming up with explanations is simply to explain, to some level of satisfaction. We came to suspect that there is much more to it. Our initial insight came when we encountered a story about the history of a famous prison that had been turned into a museum. The story was told by a former prison guard who now worked as a museum guide. He discussed a stain in the concrete floor, said to be the spilt blood of an inmate who died swearing revenge on his killer. As the prison guide explained, all attempts to remove the stain had failed. He said, in an eerie tone of voice, "No one can explain it!" This was an attempt to tell a causal story with the goal of *preventing* the listener from engaging anything but a mystical explanation of cause. The purpose of the reasoning was to influence (prevent) the causal reasoning of someone else, not just to explain something.

Our analysis uncovered additional reasons for initiating causal reasoning, such as deception and influence. Other purposes include the goal of understanding one's own actions or beliefs (which we call "ipsative" causal reasoning, from the Latin *ipse*, meaning "self"). Ipsative reasoning can be a fuzzy boundary away from metacognition, that is, one can ask questions about one's own causal reasoning, which we call *reflexive* reasoning. In the manner of abduction, one can judge the plausibility of a hypothesized future (recognition that there is an explanatory gap in one's own reasoning). One can reason about someone else's causal reasoning, which we call *projective* reasoning, that is, reasoning about what someone else thinks might happen, or about how to influence someone else's reasoning, or how to prevent someone else from engaging in correct causal reasoning (that is, deception). This would also include reasoning about what an intelligent system is doing.

As valuable as philosophical analysis might be, what forms of causal reasoning do we actually see when we adopt a naturalist method and look out at the world of human activity and explanation?

What these considerations implied is a need for a new kind of taxonomy, one that is considerably richer than the lists of "types of causes" one finds in the literature (see the article in the last issue of this magazine[9]). Table 1 presents a taxonomy of the purposes of causal reasoning.

Only a few of the causal reasoning cells in Table 1 have been subject to empirical investigation. But clearly the other forms occur widely in daily human experience.

And there is more besides. All the Table 1 entries reference what might be called *observative* causal reasoning, in which the reasoner is a commentator on or analyst of the to-be-explained events. This should be distinguished from *agentive* causal reasoning, in which the reasoner has a causal power in the events that are being explained. The possession of an agentive or causal role makes a significant difference in the shape of one's causal reasoning.

The projective reasoning column in Table 1 involves reasoning about the goals, motivations, or actions of other people. However, people are also increasingly reasoning (and worrying) about the goals, actions, and processes of complex intelligent systems, including systems that perform typical tasks associated with artificial intelligence (AI).[10,11] Some of these

Table 1. Some natural purposes of causal reasoning.

Type	Ipsative ("self") causal reasoning	Projective ("other") causal reasoning
Prospective	Reasoning about the future (forecasting)	Reasoning about what someone else thinks will happen
Interventive	Natural experiment or anecdote	Deliberate experimental action to probe the cause-effect relation or test some theory
Inspective	Comprehending the present (nowcasting)	Reasoning about what someone else thinks is happening
Retrospective	Reasoning about past events (hindcasting)	Reasoning about what someone else thinks has happened
Reflexive	Reasoning about one's own reasoning, for example, "Why is this difficult?"	Reasoning to influence someone else's reasoning, for example, deception
Continuous	When do I have an account? What is the stopping rule?	Reasoning to prevent someone else from engaging in causal reasoning
Corrective	Recognition that there is an explanatory gap Reasoning about what went wrong in one's causal reasoning When do I change my explanatory account? How do I know when to change it? Responsive gap filling (response to encountering a black swan)	Recognition that there is an explanatory gap Reasoning about what went wrong in someone else's causal reasoning Responsive gap filling (response to encountering a black swan)
Protective	Reasoning to achieve a justification or rationalization of one's actions, to provide a rationale (for example, "cover your butt")	Reasoning to achieve a justification of rationalization of someone else's (or some organization's) actions, to provide a rationale (for example, "cover your butt" and "scapegoating")

functions include optimization (for example, a GPS routing system), information retrieval (such as search engines), image classification and captioning (often with "deep" neural networks), and interaction with complex "autonomous" systems. These systems are interesting because, unlike traditional computational tools that may be subject to causal understanding, they themselves assist in performing causal reasoning. Consequently, it might be useful to consider the taxonomy in Table 1 in terms of typical reasoning roles that might epitomize domain expertise for different purposes of reasoning.

The roles identified in Table 2 are examples of specific causal reasoning activities with specific purposes.

Clearly, the roles (leftmost column in Table 2) each involve many different activities, but many have a central task that illustrates a different causal reasoning purpose. We engage in causal reasoning for many different purposes. The nature of the causal reasoning depends on these purposes.

Yet even these distinctions might be inadequate. Some of the interventive roles perform causal reasoning to determine a course of action (diagnostician, fixer, teacher), and so causal reasoning will end, or at least shift to some other form, when a course of action is determined. For example, a physician might not need to distinguish different infections if the same antibiotic is the treatment; a car mechanic might not need to determine why a muffler is loud if the solution is to replace the muffler. Other interventive roles, such as a scientist, might have less determinate purposes.

It should also be clear that for many of these roles, intelligent and data-based systems are increasingly being used to help individuals conduct their causal reasoning. Consequently, this might be thought of as creating something like "augmented causal reasoning" in which reasoners not only reason about events in their domain of their expertise or focus of concern, but their reasoning is assisted by the assistive system, and their reasoning is, at least in part, about how that assistive system works.

For example, traditional drivers might have a favorite route around a city that they take when there is traffic or construction activity. When assisted by a GPS router, it might give them a different route, which they might choose to either use or ignore, based on their understanding of whether the GPS router knows about the construction, or the traffic, or other factors the drivers may understand. Thus, the casual reasoning is both about the world (if I go on this route, I may be stuck in traffic), but also about the GPS algorithm (the GPS is telling me to go through the area with the most traffic, which is shorter, but will end up taking longer). Even if the GPS system reflects traffic conditions, experienced drivers might know that a big plant is about to end its shift, so the traffic won't appear for another 5–10 minutes, too late for the drivers to change their route.

Themes of Causal Explanation

In another study, we collected 74 newspaper and magazine articles illustrating causal reasoning with the goal of sampling varied venues of human activity including sports, politics, world events, and economics. The subprime mortgage crisis provided many explanations as the debacle unfolded. The 2007–2008 American football playoffs and Super Bowl offered different types of accounts. The Republican and Democratic primaries generated ample speculations about the reasons why different candidates succeeded and failed. The changing

Table 2. Roles and tools for causal reasoning.

Role/reasoning	Purpose	Example assistive tools
Diagnostician (physician, mechanic)	Choose cause to determine treatment (*interventive*)	Computer diagnosis models, sensors, diagnostic tests
Detective	Identify and/or punish culprit (*retrospective*)	DNA/fingerprint matches, large database queries to identify potential culprits
Entrepreneur	Identify underserved market, products, or customers that will lead to profit (*prospective*)	Demographic models, market segmentation, and market research models
Daytrader	Predict shifts in a stock price (*continuous, prospective*)	Complex predictive financial models
Weather forecaster	Forecasting weather so we can plan a day better (*continuous*, *prospective*)	Complex weather, climate, and atmosphere models
Accident investigator	Identify reason for accident and assign blame (*protective, projective*)	Reconstructive models from black-box and other data
Air traffic controller	Maintain awareness of complex environment to maintain safety and route aircraft (*inspective*, *prospective*)	Situational awareness systems that identify where/who/what are in the airspace and predict future states
Sports bookmaker	Identify initial odds/point spread; adjust based on betting to set fair or advantageous odds (*prospective*)	Bettor model; rating percentage index
Tax auditor Forensic accountant	Identify whether reported income is accurate (*retrospective*, *projective*)	Forensic models identifying key predictors of fraud
Speculator or prospector	Tie up resources on the hope that some will pay off (*prospective*)	Domain models that predict future payoffs (geological models, economic models)
Experimental scientist	Test hypotheses using empirical methods to identify "truth" (*interventive*)	Inferential statistics
Epidemiologist	Test hypotheses in archived data to identify health patterns (*retrospective*)	Inferential statistics
Defense attorney	Establish alternative causal theory that pertains to a case (*protective*, *retrospective*)	Case law search queries
Autonomous rider/ GPS-guided driver	Identify whether a routing system is behaving properly (*continuous*)	Optimization of complex cost functions (for example, Dijkstra's algorithm)
Student	Identify how to learn a curriculum (*reflexive, ipsative*)	Automated tutors to help understand knowledge and where more instruction is needed
Teacher	Identify how to teach a curriculum *(reflexive, projective, interventive*)	Deciding which achievement tests to use to model student progress and determine where help is needed
Sportscaster	Create narrative to explain story of game/series/season (*retrospective, inspective*)	Statistics and analytics

conditions in Iraq stimulated analyses of what went right and wrong.

In each of the articles, we identified the individual statements of causal attribution, we labeled the statements with identifiers, and we made notes that summarized each attribution. For instance, one story offered an explanation for the increasing cost of products made in China (the effect X to be explained). Some causes led directly to the effect, as in a "chain." For example, China reduced and removed tax incentives for exporters of Chinese goods (A), which led to increased costs of exports (A → X). Product recalls and environmental crackdowns (B) also led to increased cost of products made in China (B → X).

Causes were also indirect. For example, an increase in oil costs (C) led to an increase in the cost of plastics (D), which led to an increase in the cost of Chinese products (C → D → X). Labor shortages and stricter labor rules (E) led to an increase in wages, which (F) led to an increase in the cost of Chinese products (E → F → X). This seemed to be a "swarm" of converging effects, but it had some chains of effects within it.

As we collected and analyzed more accounts, we began to see some themes to the structures of the causal explanations. In some explanations, the cause was seen as a single dramatic event that could have gone the other way (for example, a basketball team lost a game because of a basket at the very end of a game), whereas in others there was a critical event but it was not so dramatic, coming earlier in the event sequence. Cases of these types suggested a theme we call "the reversible," a single condition for how something could happen (for example, HIV causes AIDS). We also found stories in which the mechanism was complex, involving multiple causes in which the effects interacted with one another.

The incident accounts often referenced more than one cause; we tallied 219 individual causes. Only two of the 39 sports incidents referenced 10 or more causes. Four of the 18 economics incidents included 10 or more causes. None of the political, military, or miscellaneous incidents had even 10 causes. That said, causes are often bundled together. We identified three common ways for them to be bundled into a higher-level explanation.

Events are mutable, that is, reversible events, actions, or decisions, commonly referred to as *counterfactuals*. For example, late in the last quarter of the 2008 Super Bowl between the New York Giants and the New England Patriots, Eli Manning, the Giants' quarterback, seemed almost sure to be sacked by the opposing Patriots but somehow spun away and got off a pass that the receiver caught against his helmet. Most accounts of the game highlighted this miracle play because if Manning had been sacked the game would almost certainly have ended with the Giants losing, and it was very easy to imagine the play failing if Eli were sacked, as he appeared to be. As we analyzed more accounts, we came to expect that sports incidents often invoke reversal (counterfactual) explanations, but reversibles were expressed for events in other domains. For example, in the economics category, the US Federal Reserve decision to keep interest rates low in the period 2002–2004 has been identified as a cause of the housing boom, the housing bubble, and the subsequent recession.

*Abstraction*s take several causes, sometimes including counterfactuals, and synthesize them into a single explanation. In professional US basketball, a series of mistakes by the New York Knicks was synthesized to explain why the Knicks lost the game—too many mistakes. The abstraction theme was more prevalent for sports than for economics. The abstraction is sometimes offered by itself, with exemplars being implicit, but at other times the abstraction was used as a way to bundle events in which all of the relevant factors and events are of the same kind. Most important, an abstraction is usually offered as a single answer to the question of what caused an event, in contrast to lists and stories. Abstractions are not always simplistic, however, because they can blend a set of individual causes that share common features.

Conditions are in effect even before the to-be-explained event began. Thus, in sports, if a key player was so injured that he did not even play, we counted that as a condition because it did not occur during the contest. Economics offers many examples of conditional explanations—a market force inexorably at work, such as the development and collapse of bubbles. Often, a condition theme is used in a simplistic fashion. The economic recession is blamed on greed. The success of a sports team is attributed to better coaching, or the fact that they "wanted it more." Or consider the cause of World War I. The assassination at Sarajevo explains it as an event, whereas the rise of nationalism explains it as a condition—a feature of the situation.

A *list* is merely multiple reasons why something happened. Lists are fairly common in explanations of sports outcomes—for example, the reasons the Patriots lost the Super Bowl. For the sports category, 14 of the 38 accounts featured a list. Lists are less common in economics, although an example would be an article listing the reasons why the Chinese economy should move to a higher rate of inflation. All of the articles on politics relied on a list—the reasons the political campaigns of John Edwards, Rudy Giuliani, Mitt Romney, or Hillary Clinton (in 2008) folded.

Stories provide a deeper analysis to present a mechanism of how multiple causes interacted. Sometimes stories took the form of a chain. Chains were relatively rare in the sports incidents, and when they were used, they were very short. Chain-reaction stories seem more prevalent in economics. In general, economics analyses used the most complex story explanations, that is, they are not always chains. For example, one article described how the Federal Reserve worsened the subprime mortgage problem. It described the interaction of multiple, parallel causes (interest rates, inflation, the housing market, oil exports, and so on). An article explaining the death by asphyxiation of a fireground commander in New York presented multiple reversibles formed as a chain. One set of reversibles related to the spread of the fire into the hall, and another set referred to the failure of the lieutenant to withdraw in time. So this case was neither a single event nor a simple chain.

Most explanations of events in the economics domain did not include any counter-causes, that is, countervailing forces or opportunities for events to unfold differently. Economic events are perceived to be strongly determined. In contrast, many of the sports accounts note counter-causes. A few of the Super Bowl accounts noted that the Giants were lucky with their miracle play, which changed the outcome of the game. Of the 38 sports incidents, 12 cited some sort of counter-cause. Only three of the 18 economics incidents did so. Sports accounts seem to be more sensitive to factors such as luck, and sometimes offer a counterfactual perspective

that is usually missing from analyses in economics.

In sum, what we have found is that it is possible to identify certain kinds of explanations that seem more common—people seem to find them more useful or comfortable—in explaining different kinds of events.

Causal Explanation Templates

Using concept map-like diagrams, we have been able to express the basic structure of a number of kinds of causal explanation themes. In addition to the "abstraction" and the "chain," we found a dozen themes, including the "swarm"—when multiple and *independent* causes converge or combine to cause an effect; the "clockwork"—when multiple and *interacting* causes combine to bring about an effect; the "culprit"—when one of a number of possible causes gets singled out as the cause; and the "snark"—when one wonders if one is not only looking for the wrong causes but is also looking at the wrong effect. Figures 1 and 2 present basic conceptual models of the structures for the abstraction and the clockwork.

To exemplify these models, an instance of the clockwork was in one of the economics articles: the deregulation of banks permitted mortgage building and included relaxed lending criteria; the relaxed lending criteria led to risky loans, which was the target for the mortgage building; and the risky loans led to a drop in the housing market and along with the mortgage building caused bank defaults.

While these themes, and combinations of them, described the variety of causal explanations, they do not capture the process by which the explanations are arrived at. For this, we formed a different sort of model.

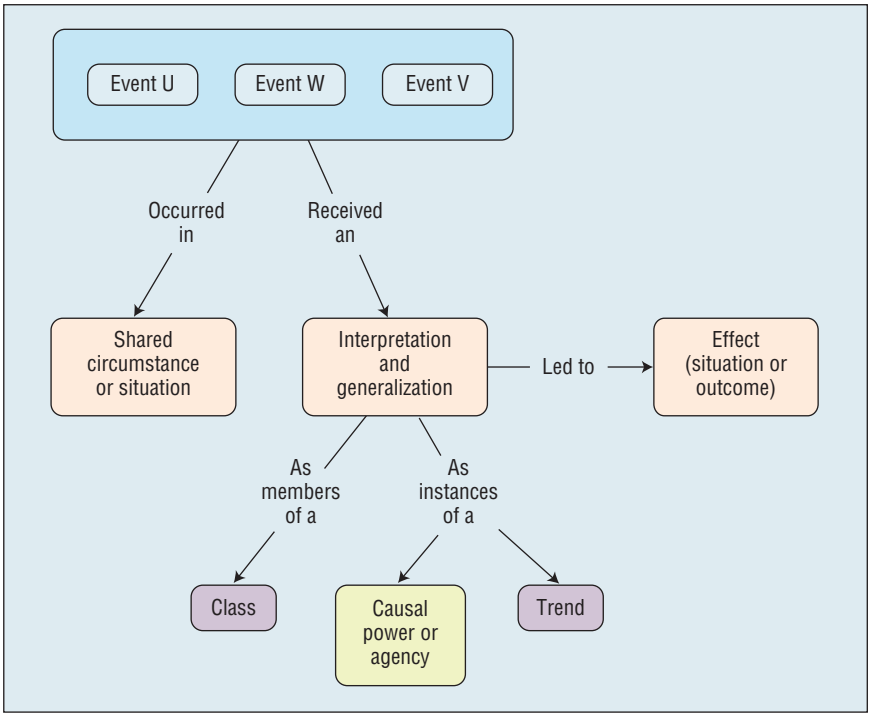

Figure 1. Conceptual model of the structure of the abstraction explanatory theme.

Process Model

Our findings suggest that when reaching for a causal explanation, people typically first make a list of possible causal events, decisions, conditions, or abstractions. The list entries are then sometimes ordered; sometimes they are synthesized into abstractions, or connected as stories. The final preferred explanation might reference a single cause (an event, a decision, a force, or an abstraction) or multiple causes (such as a chain, a clockwork, and so on). People do not always prefer the most complex explanations, such as the clockworks that show multiple interactions of causal variables. Our results suggest that as people begin the process of causal sensemaking, they initially latch on to simple, single causes (sometimes reversibles and sometimes abstractions), then they expand and deepen the analysis, but only as little as is necessary. The tradeoff is that the simpler explanation is easier to communicate, to remember, and to use as a basis of projection to the future. It did not escape our notice that his pattern characterized our own struggles in trying to develop a robust and broadly applicable taxonomy for causal reasoning. We would see instances that suggested a type of cause or a theme, and build upon the list of themes, but then we would feel a need to reduce and simplify.

> Sports accounts seem to be more sensitive to factors such as luck, and sometimes offer a counterfactual perspective that is usually missing from analyses in economics.

Our current model of causal explanation was driven by our main

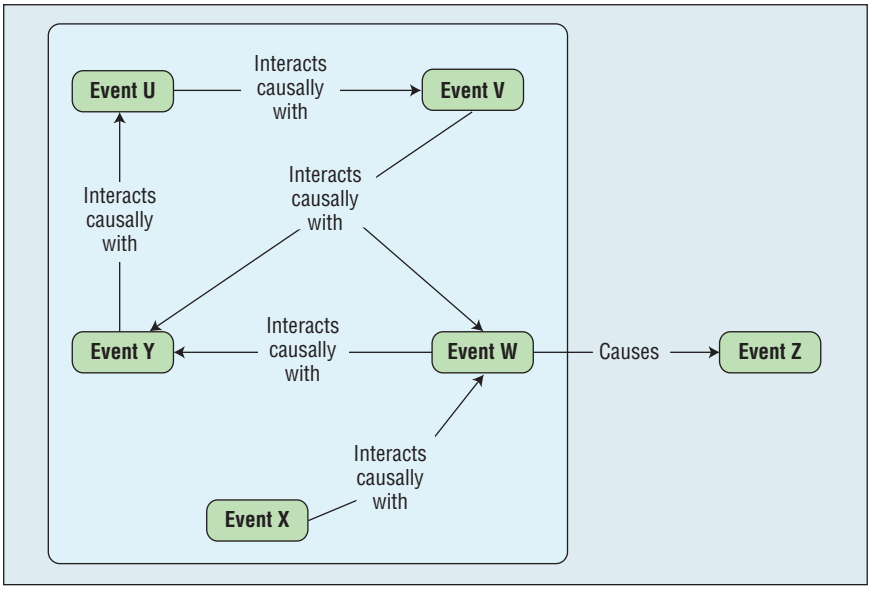

Figure 2. Conceptual model of the structure of the clockwork explanatory theme.

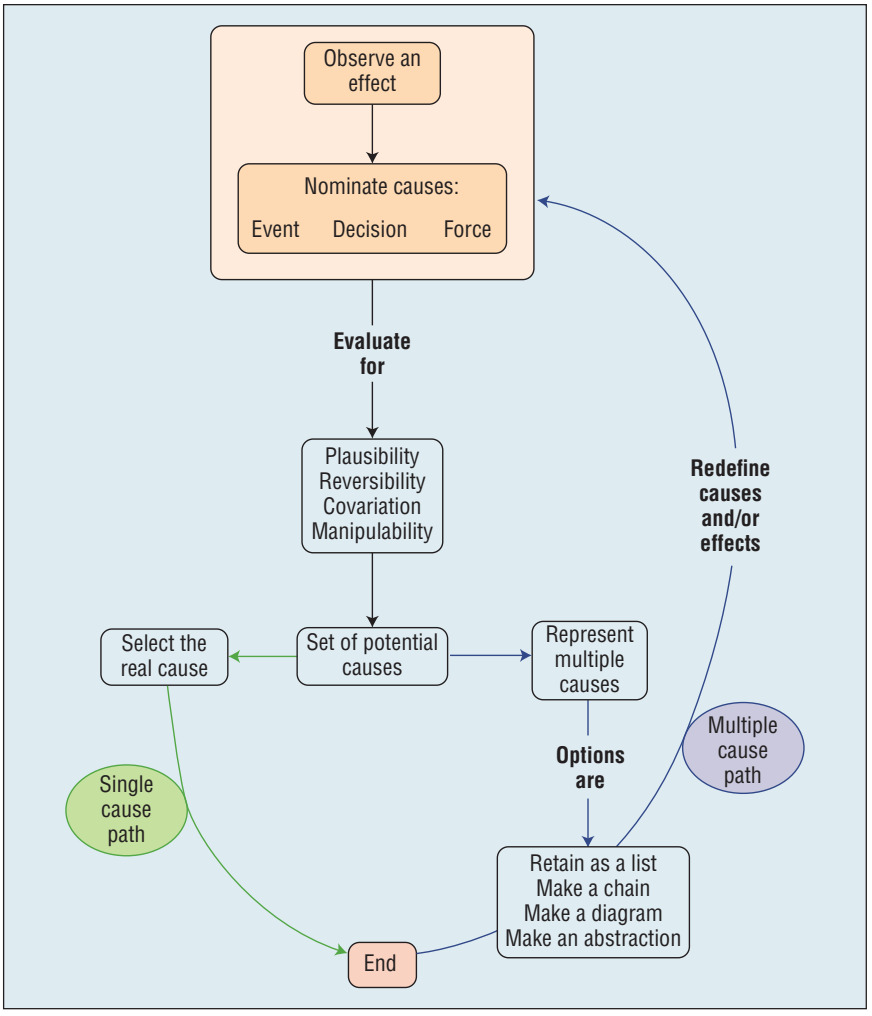

Figure 3. A macrocognitive model of indeterminate causal explanation.

consideration that causal reasoning, as a form of sensemaking, should conform to the general structure provided in the data/frame theory.[12–14] The data/frame theory of sensemaking asserts that explanations are frames (or mental models) that get built out of data, while at the same time, what counts as data is defined by the frames that are being used to fashion explanations. If we apply this perspective to the formation of causal explanations, the causes identified in a situation (based on propensity, covariation, and reversibility) generate the explanatory frames (as illustrated in Figures 1 and 2). At the same time, these explanatory frames (events, abstractions, conditions, lists, and stories) guide the search for causes and a most satisfying explanation. Both processes occur simultaneously.

For example, physicians in the United States and Europe who were detecting the onset of AIDS were noticing coincidences across their patients, but these coincidences were not simply a matter of matching patterns because the features hadn't been discovered before, and each case showed different symptoms (because AIDS is an opportunistic infection). Rather, the detection of coincidences was conditioned by the types of mental models and explanations that the physicians had learned.

Figure 3 presents a macrocognitive model of causal reasoning based on our findings. This can be understood as a merger of the data/frame notion with steps 1–4 in Charles Sanders Peirce's model of abductive inference, presented in the previous article on causal reasoning in this department.[9]

What "Counts" as an Explanation?

In another of our experiments, a large sample of American and ethnic

Table 3. Myths about causal explanation.

Myth	Reality
Correlation does not imply causation.	Of course it does. It does not require a determination of causation, but it is often the beginning of a fruitful investigation.
Logic is the exclusive basis for the analysis of causal reasoning.	Perhaps in philosophical investigations, but in real-world settings the evidence for causation is usually too ambiguous to permit valid deductive inferences.
The analysis of physical causation is the model for understanding all forms of causation.	Of all the events about which humans reason, speculation about causation more often involves indeterminate questions for which there will never be any closure on the single or "real" cause.
The scientist is the ideal model for causal reasoning.	Much of the research on causal reasoning involves scientists learning to overcome reductive tendencies to oversimplify cause-effect relationships. However, in natural settings some degree of simplification is necessary to cope with complexity, and furthermore, scientific standards are too restrictive.
Causal reasoning means finding the one true cause for an effect.	This might be true for studies of physical causation, but it is not true for natural settings involving indeterminate causes. The quest for a single "root" cause must be a distortion and an oversimplification, although people often seek single causes as a desirable simplification.
The "effect" to be explained is usually clear.	In natural settings, people often revise the description of the effect as the causal investigation continues.
Causal reasoning is to be described as a process having clear-cut beginnings and endings.	Quite often, it does not.
The property of "being an explanation" is a property of statements.	Clearly, it is a complex interaction.

Chinese college student participants were given two scenarios involving indeterminate causation, one regarding the US economic collapse in 2007–2008 and the other about the US military engagement in Iraq in 2007.[15] Participants were asked to assess three types of causal explanations for each of the scenarios: single causes expressed as single sentences, lists of multiple causes, a causal chain explanation, and a causal network diagram. Participants were asked these questions:

- Which of these types of explanations are the most satisfying to you?
- Which of these types of explanations would you give to your 12-year-old nephew who wants to understand what happened?
- If a newly elected politician was put on an action committee, which of these explanations do you think that politician would prefer?

The participants preferred the simple sentence explanations and the list format over the diagrams when the intended recipient of the explanation was a young nephew, but for a politician or themselves, they preferred the list and diagram format over the simple sentences. Ethnic Chinese participants preferred complex explanations more than the Americans. The form of presentation made a difference: participants preferred complex to simple explanations when given a chance to compare the two, but the preference for simple explanations increased when there was no chance for comparison, and the difference between Americans and Chinese disappeared.

In sum, we were able to deliberately manipulate "what counts" as a preferred explanation. Referencing Figure 3, we could shift participants from the single-cause path at the left (the simple explanations) to the multiple-cause path at the right (the complex explanations). In addition, we demonstrated what appears to be an individual difference in preferences that relate to culture.[16] Preferences for explanations can vary with the context and with the audience, and they depend on the nature of the alternatives that are provided.

> *The property of "being an explanation" is not a property of statements: It is an interaction of statements with knowledge, context, and intention.*

This has profound implications for intelligent systems that are intended to generate explanations or to help people generate explanations.

A widely held belief is that in causal reasoning people identify an effect, nominate causes, and select what they believe is the best one. And then the process is over. This approach fits some contexts such as scientific investigations. It does not always fit causal reasoning. More typically, the initial effect may be reframed and recast during the investigation into its causes.

In our research described here, we identified a number of mistaken beliefs about causal reasoning, which we list in Table 3.

The taxonomy and models we have presented make it clear that causal explanation is rich with interesting avenues for extending our empirical base on the unexplored varieties and forms of explanatory-causal reasoning. The third essay in this series describes a "causal landscape" as a method that would enable decision makers to engage in causal reasoning that leads to actionable conclusions. It is also suggestive of a way of coping with complexity. As such, it might empower the developers of intelligent systems to go from their complex and abstract understandings of their systems to succinct explanations that would enable users to develop justified trust in the technology and take correct actions in the use of the technology. ■

References

1. A. Gopnik and L. Schulz, eds., *Causal Learning: Psychology, Philosophy, and Computation,* Oxford Univ. Press, 2007.
2. S. Sloman, *Causal Models: How People Think about the World and Its Alternatives*, Oxford Univ. Press, 2005.
3. J. Reason, *Human Error*, Cambridge Univ. Press, 1990.
4. P.J. Feltovich, R.R. Hoffman, and D.D. Woods, "Keeping It Too Simple: How the Reductive Tendency Affects Cognitive Engineering," *IEEE Intelligent Systems*, vol. 19, no. 3, 2004, pp. 90–95.
5. D. Dörner, *The Logic of Failure: Why Things Go Wrong and What We Can Do to Make Them Right*, Perseus, 1996 (original work published 1989).
6. D. Hume, *A Treatise of Human Nature*, anonymous publisher, 1739–1740; reprinted by New Vision Publications, 2007.
7. D. Kahneman and C.A. Varey, "Propensities and Counterfactuals: The Loser That Almost Won," *J. Personality and Social Psychology*, vol. 59, no. 6, 1990, pp. 1101–1110.
8. R.R. Hoffman, G. Klein, and J.E. Miller, "Naturalistic Investigations and Models of Reasoning about Complex Indeterminate Causation," *Information and Knowledge Systems Management*, vol. 10, nos. 1–4, 2011, pp. 397–425.
9. R.R. Hoffman and G. Klein, "Explaining Explanation, Part 1: Theoretical Foundations," *IEEE Intelligent Systems*, vol. 32, no. 3, 2017, pp. 68–73.
10. M.T. Ribiero, S. Singh, and C. Guestrin, "'Why Should I Trust You?' Explaining the Predictions of Any Classifier," *Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining* (KDD 16), 2016, pp. 881–833.
11. M.D. Zeller and R. Fergus, "Visualizing and Understanding Convolutional Networks," *Computer Vision—ECCV 2014*, D. Fleet et al., eds., LNCS 8689, Springer, 2013, pp. 818–833.
12. G. Klein, B. Moon, and R.R. Hoffman, "Making Sense of Sensemaking 1: Alternative Perspectives," *IEEE Intelligent Systems*, vol. 21, no. 4, 2006, pp. 22–26.
13. G. Klein, B. Moon, and R.R. Hoffman, "Making Sense of Sensemaking 2: A Macrocognitive Model," *IEEE Intelligent Systems*, vol. 21, no. 6, 2006, pp. 88–92.
14. G. Klein et al., "A Data–Frame Theory of Sensemaking," *Expertise out of Context: Proc. 6th Int'l Conf. Naturalistic Decision Making*, RR. Hoffman, ed., 2007. pp. 113–158.
15. G. Klein et al., "Influencing Preferences for Different Types of Causal Explanation for Complex Events," *Human Factors,* vol. 56, no. 8, 2014, pp. 1380–1400.
16. M. Morris and K. Peng, "Culture and Cause: American and Chinese Attributions for Social and Physical Events," *J. Personality & Social Psychology*, vol. 67, no. 6, 1994, pp. 949–971.

Robert R. Hoffman is a senior research scientist at the Institute for Human and Machine Cognition. His research interests include macrocognition and complex cognitive systems. Hoffman has PhD in experimental psychology from the University of Cincinnati. He is a Fellow of the Association for Psychological Science and the Human Factors and Ergonomics Society and a Senior Member of IEEE. Contact him at rhoffman@ihmc.us.

Shane T. Mueller is an associate professor of psychology, cognitive, and learning sciences at Michigan Technological University. His primary interest is implementing formal quantitative mathematical or computational models of memory, perception and decision making. Mueller has a PhD in psychology from the University of Michigan. Contact him at shanem@mtu.edu.

Gary Klein is chief scientist at MacroCognition, LLC. His research interests include naturalistic decision making. He is a Fellow of the American Psychological Association and the Human Factors and Ergonomics Society. Contact him at gary@macrocognition.com.