Fodor, J. (1987). Modules, frames, fridgeons, sleeping dogs, and the music of the spheres. In Z. W. Pylyshyn (Ed.), *The robot's dilemma: The frame problem in artificial intelligence* (pp. 139–149). Norwood, NJ: Ablex.

Hayes, P. J. (1973). The frame problem and related problems in artificial intelligence. In A. Elithorne & D. Jones (Eds.), *Artificial and human thinking* (pp. 45–59). Amsterdam: Elsevier.

Hayes, P. J. (1987). What the frame problem is and isn't. In Z. W. Pylyshyn (Ed.), *The robot's dilemma: The frame problem in artificial intelligence* (pp. 123–137). Norwood, NJ: Ablex.

Hofstadter, D. (1995). *Fluid concepts and creative analogies*. New York: Basic Books.

Maunsell, J. H. R., & Newsome, W. T. (1987). Visual processing in monkey extrastriate cortex. *Annual Review of Neuroscience, 10*, 363–401.

McCarthy, J., & Hayes, P. J. (1969). Some philosophical problems considered from the standpoint of artificial intelligence. In B. Meltzer & D. Michie (Eds.), *Machine intelligence* (Vol. 4, pp. 463–502). New York: Elsevier.

McDermott, D. (1987). We've been framed: Or, Why AI is innocent of the frame problem. In Z. W. Pylyshyn (Ed.), *The robot's dilemma: The frame problem in artificial intelligence* (pp. 113–122). Norwood, NJ: Ablex.

Pylyshyn, Z. W. (Ed.). (1987). *The robot's dilemma: The frame problem in artificial intelligence*. Norwood, NJ: Ablex.

Raphael, B. (1976). The frame problem in problem-solving systems. In N. Findler & B. Metzger (Eds.), *Artificial intelligence and heuristic programming* (pp. 159–169). Edinburgh: Edinburgh University Press.

Roediger, H., & Craik, F. (Eds.). (1989). *Varieties of memory and consciousness: Essays in honor of Endal Tulving*. Hillsdale, NJ: Erlbaum.

Rogers, H. (1987). *Theory of recursive functions and effective computability*. Cambridge, MA: MIT Press.

Sloman, A. (1989). On designing a visual system. *Journal of Experimental and Theoretical Artificial Intelligence, 1*(4), 289–337.

Tulving, E., & Schacter, D. (1990). Priming and human memory systems. *Science, 247*, 301–306.

CHAPTER 3

The Adventures Among the Asteroids of Angela Android, Series 8400XF with an Afterword on Planning, Prediction, Learning, the Frame Problem, and a Few Other Subjects

Clark Glymour
Department of Philosophy
Carnegie Mellon University

Angela's attention drifted as the data on the changing physical conditions in the vicinity of Bernard IV flashed by on the computer screen. She longed for her days as a nuclear reactor robot, or Nuke Bot as they were known, where a genitally and hormonally correct series 8400 F-type android such as she could find a little action after work, have a good time with a brawny droid, maybe even with a guy with half a brain. Rotten luck—she'd been cooked when the reactor leaked. Too radioactive for Earth work, reprogrammed with the latest causal prediction system, here she was in a rocket with Shiela, a series 8300 F type, uncooked, heading for Bernard's star. All of the male company was in other ships on the expedition; alpha 1 had the two series 8300 M type nonmonotonic reasoners, one with good pecs; alpha 2 had the fellows with the ID2000 classification and prediction programs. Like Harry and Felix in alpha 1, Shiela had the standard heuristic learning programming, nonmonotonic reasoners, and built-in belief networks, but Angela, since reprogramming, was an X type, and she had authority to override Shiela's decisions. The three ships comprised the first android expedition to another star system; much too far for humans, who were weenies anyway, Angela thought.

Remembering the good old days wasn't going to help now. Something was clearly wrong around Bernard IV, which looked to be the only habitable satellite of Bernard's star. The planet was surrounded in all directions by asteroids. In principle there was no reason why the asteroids should prevent the ships from coming close to Bernard IV, or even landing. The asteroid motions should be predictable, and the expedition ships could just avoid them. But these asteroids weren't behaving correctly; every now and then one of them would change direction and swerve out of its computed trajectory. The things behaved as though they had some sort of propulsion system. How could Angela keep the damned things from swerving into their ships? Unless she could predict the swerve—technically the non-Newtonian acceleration—or figure out how to control it, she couldn't avoid a collision, and she'd never get to have another beer with a good-looking droid. She had ordered the ship to measure every physical variable correlated with the swerve, or correlated with any variable correlated with the swerve, and now the numbers were running by.

"Pool it and compute correlations, Shiela. Then discretize it and compute a contingency table. I'm gonna call alpha 1." She punched in the communication code.

"Harry, Felix, this is Angela calling. You computing?"

"Yes, Angela, we are computing," Harry's voice responded.

"We have a problem. Bernard IV is surrounded by asteroids."

"Yes, we know, Angela."

"Asteroids move in Newtonian orbits, Harry."

"Of course, Angela."

"The data say these asteroids don't move in Newtonian orbits, Harry."

"Then either they aren't asteroids, or else the data are in error, Angela."

"Thanks for nothing, Harry. What the fuck should we do so the asteroids don't hit us?"

"Angela, please don't be rude. We will have to calculate the minimal perturbations of our present beliefs that regain consistency. The problem will require some considerable computing."

"Yeah sure, Harry. By the way, Harry, Tweety is a bird. Can Tweety fly?"

"Yes, Angela, since Tweety is a bird Tweety can fly."

"But Tweety is an ostrich, Harry."

"Oh, then Tweety can't fly, Angela. I will have to recompute."

"Thanks, Harry. Out."

Shiela frowned at her. "You shouldn't tease him like that Angela." "Can't help myself. Those nonmonotonic droids are such feebs." "I'm standard nonmonotonic, Angela." "Yes, but you're F type and you're my partner, so it doesn't apply. What do the correlations look like?" "Well, the variables kicked out are pretty strange. Using your criteria the system picked out the non-Newtonian acceleration, proximity of the asteroid to Bernard IV, the mass density of two spectral types of

rocks within a small radius from the asteroid—roughly red rocks and green rocks—the mass density gradient in the region surrounding the asteroid, the density of dust in storms on Bernard IV, and the intensity of monochromatic blue flashes coming from Bernard IV. It looks like a mess to me Angela."

"Put up the correlations, Shiela." Shiela punched up the display.

Correlation Coefficients
N = 2,014

p	s	f	a	m	d	g
1.00000						
0.5493	1.00000					
0.25055	–0.03528	1.00000				
0.36976	0.20483	0.49932	1.00000			
0.01031	–0.00743	–0.03363	–0.33786	1.00000		
–0.02820	–0.03310	–0.03150	–0.13599	0.37453	1.00000	
0.00349	0.00216	–0.00479	–0.15951	0.49679	– 0.02541	1.00000

p = proximity; s = storm intensity; f = blue flash intensity;
a = asteroid non-Newtonian acceleration;
m = mass gradient near asteroid surface;
d = density of red rocks in region of asteroid surface;
g = density of green rocks in region of asteroid surface.

Not bad, Angela thought. Two thousand observations already, but they were only going to get to do one experiment and they'd be smithereens if they predicted the outcome wrong. "Amazing," she said aloud.

"I've already sent the data and the correlations to the other ships," Shiela said. "Maybe one of them will know what to do."

"Shiela, I know what to do just looking at the correlations. Get as much mass as you can hooked up to low-velocity guidable rockets and get them ready to fire. I'm going to write a little program that will start shooting off those rockets as we enter the asteroid belt."

"Are you going to try to blow up the asteroids, Angela?"

"Not a chance, pal. I'm going to tempt the asteroids to get out of our way. Sort of a diversion."

"Don't you think, Angela, that we should consult with the other ships?"

"There is just no point talking with your friends in alpha 1. They don't have a built in belief network for this case, and they can't build a new one that could possibly be right; if I give them instructions, it will conflict with what they already believe, and they'll be computing until they're dust, or else they'll run a PUPS routine, or they'll chunk on noisy data, or some such thing. They're goners, Shiela."

"But, Angela . . . "

"Tweety is a bird, Shiela. Can Tweety fly?"

"Yes, Tweety can fly."

"But Tweety is an ostrich, Shiela."

"Oh. Then Tweety can't fly, I have to recompute."

"See what I mean Shiela? You've got the logical form, the semantics and all that, but you don't know shit from Shinola, and with your stupid heuristic learner you can't learn it unless somebody with authority tells you."

"What's Shinola?"

"Nobody knows. It's just a saying. Please get to work with the rockets. And as soon as we approach the asteroid belt I want you to beam music and friendly talk toward Bernard IV on as many radio frequencies as possible."

"Why, Angela?"

"Because something is punching the asteroids away from Bernard IV. Whatever it is, it probably comes from the planet. I don't know if it just hits asteroids or if it hits any massive body approaching Bernard IV. I don't want it to hit us. Since unusual phenomena associated with planets may be due to life forms, we want to give anything down there every opportunity to recognize that we're no threat. Get to work."

Angela punched in the alpha 2 code. "Basil, this is Angela. What's your take on the asteroid problem?"

"We've analyzed it, sweetheart, and we've got the solution. Everything says the key is the blue light. Flash it and the asteroids swerve—away, not toward, the source. The blue light is far and away the best predictor of the swerve. So we've put a filter on our beams. Whenever one of those things gets close, we'll flash the light. You do the same."

"What about alpha 1, Basil?"

"They're dinks, still trying to recompute and figure out what to do with their primitive learner. You know they're hopeless. What do they run? Some descendant of PUPS or some deterministic chunker? I don't have authority over them. They're dead silicon."

"You will be too, Boris, if you rely on flashing a blue light."

"What do you mean, Angela?"

"The light has no effect on the asteroids, Basil. It's caused by some force we can't see, probably from Bernard IV. Wherever and whatever it is, it also causes the asteroids to swerve away from the plant."

"Be serious, Angela, how can you know that there's a force you can't see? What is it, the tooth fairy force? And even if you were right, how are you going to avoid the asteroids?"

"I know it from the data Basil. I'll tell you how when you've got the time. Right now you've got to start loading your guidable rockets with as much mass as possible. In a few minutes I'll send you a program to control them."

"Why rockets with mass? You think the non-Newtonian acceleration is caused by the red and green rocks, Angela? I haven't got enough red and green stuff to matter."

"Red and green don't have anything to do with it, Basil. Differences in proximate mass density cause the asteroids to follow the gradient. Control the mass density close to the asteroids and you'll control their acceleration."

"Angela this is nonsense. The correlation between blue flashes and accelerations is something like .86, and the correlation between mass density differential and swerves is only –.4, but you say you know that the flash isn't a cause of the acceleration and the mass density differential is? And you know the color doesn't matter, even though the red and green rocks are correlated with the acceleration? I thought droids weren't sensitive to drugs. You must have a defect, honey. Everything in statistics and the *Handbook of Android Epistemology* says you're wrong."

"Please Basil, don't confuse correlation with causation. I do know because of my experimental program. Assemble the rockets."

"Me confuse correlation and causation? Angela, honey, your confusing causation with weak correlation. Go build a blue filter."

"Basil, you overpriced regression package, you bastard son of misspent taxpayer dollars, you don't know shit from Shinola."

"What's Shinola, Angela?"

"Never mind. Out."

As alpha 3 entered the asteroid belt, the guidable rockets fired and passed near to any approaching asteroid. The asteroids followed the rockets like fish after bait. Angela and Shiela watched appalled as alpha 2 flashed its blue light at an oncoming asteroid and was pulverized for the trouble. Alpha 1, apparently following no coherent strategy, lasted a bit longer, but soon it too smashed into an asteroid that seemed to be seeking it.

Alone beneath the asteroid belt, alpha 3 settled to the surface of Bernard IV. Angela and Shiela emerged to find themselves surrounded by strange creatures, who seemed by their manner and apparatus very intelligent and very advanced. After a bit of fooling around with gestures, artifacts, and sounds, one of the Bernards began to speak perfect 21st-century English.

"We are very glad you arrived safely. Over the millennia we have seen many vehicles try to pass through what you call the asteroid belt, and all of them perished, just like your companion vehicles. You were very fortunate, but you will never have such good fortune again. You must remain here. You would surely die trying to pass through the belt again."

"It wasn't luck at all, Bernie," said Angela. "It was good planning. The asteroids aren't just rocks are they? They're sort of living creatures, right?"

"Indeed, of a very primitive kind. They are attracted by very proximate mass, which they consume and convert to energy to propel themselves."

"And so now and then they swerve out of a stable orbit and head toward a collision with your planet, right Bernie? And when one of them gets anywhere near the outer atmosphere of Bernard IV it generally kicks up a storm, right? And if they should hit the planet very often they'd ruin your atmosphere and destroy civilization?"

"How astute of you, Angela. Striking the planet would kill the asteroid, but it would do enormous damage to us. It's happened very rarely."

"Do you guys have some sort of repeller beam that gives off an intense blue flash when you fire it? I figure you fire the repeller beam to divert any approaching asteroid. The harder you shoot the more intense the flash. But sometimes the flash is obscured by the storm the asteroids kick up, right?"

"This is amazing, Angela. How could you know so much about us? Are you psychic? We thought we were the most intelligent of races, but your inferences astound us."

"It's all in the data, Bernie. Plus a good algorithm and a little common sense. I wasn't sure you guys were here, but it looked to be a likely explanation of the data."

"You must teach us your methods of divining, Angela."

"Sure thing. I'll give you the book on it. But tell me, since we're going to be visiting a while, if Tweety is a bird then Tweety can fly, right?"

"Of course, Angela."

"But what if Tweety is an ostrich?"

"Then from my understanding of your language, Tweety can't fly, Angela. So what?"

Angela looked relieved. "Bernie, I hope you guys make beer on this planet. I'm really glad to meet somebody that knows shit from Shinola."

"What is Shinola, Angela?"

AFTERWORD

The story cheats in some ways. One obvious cheat is the conversation. We do not know how to make computers talk that way. Perhaps another cheat is that Angela wrote a program to schedule rocket launches to influence the mass gradient near the surface of a rapidly moving object. The calculations are not preposterous, but I do not know enough of the state of automatic programming to know whether it is feasible for an android to write the code, even with prepackaged help.

One thing that is not a cheat is the success of Angela's inferences about causes and of her predictions about the effects of alternative courses of action. The data in the table were generated by Monte Carlo methods from a linear structure with the following equations:

$$
\begin{aligned}
p &= e1 \\
s &= 0.700p + e2 \\
d &= e6 \\
g &= e7 \\
R &= 0.0600p + e8 \\
f &= -0.300s + 0.0900R + e3 \\
m &= 0.500d + 0.600g + e5 \\
a &= -0.400m + 0.990R + e4
\end{aligned}
$$

The *e* terms are normally and independently distributed variables with unit variance; *R* is an unmeasured variable representing the generation of a repeller force on Bernard IV. With one qualification, Angela's conclusions about the casual structure and her predictions about the effects of changing the mass density or flashing a blue light can all be obtained by applying existing TETRAD II algorithms to the data (see Spirtes, Scheines, Glymour, & Meek, 1994). The qualification is that Angela's worry that the repeller force might be applied to her ship requires her to use substantive knowledge (asteroids approaching the planet are massive bodies approaching the plant; alpha 3 is a massive body), some simple reasoning techniques, and reverse inheritance (if *A* causes *B*s to *D*, and all *B*s are *C*s, maybe *A* causes *C*s to *D*).

Whatever was originally intended by the frame problem, the title quickly became associated with the general problem of computationally feasible planning of the effects of actions intended to achieve some desired goal. Prediction is essential to planning. In particular, whatever the intent of its authors, "the frame problem" became a phrase for asking how a computational agent could reliably predict the effects of its own efforts to move things. But these restrictions are entirely artificial, more artificial than the intelligence they are supposed to be about. Although all action takes place by moving something, the intermediate effects of motion may be described in other terms. Rather than asking about the effect of moving the paint brush, an android can think about the effect of causing something to change color from blue to white. Rather than thinking about the effect of moving a vibrating reed, the android can think about the effect of saying a particular thing. Rather than thinking about the effects of its own actions, in planning an android can think about the effects of someone else's actions, or even about the effects of an event that no one does, that simply happens. Any attempt to solve these problems by cutting off special cases as separate domains of expertise and research will of necessity result in solutions that are ad hoc unreliable and suboptimal.

The issues aroused by the frame problem are questions about the prediction of the effects of interventions in a causal system. To address these questions in a general way requires representing causal structures generally, as well as understanding how the representations can be used to compute such predictions. Only an evasive mind can then avoid the further question: How can an android *learn* the relevant representations of causal structures? The central questions about planning and the frame problem are: (a) How can causal structure be learned reliably, efficiently and feasibly? and (b) How can complete or partial knowledge of causal structure be used to reliably and feasibly predict the effects of interventions in the causal system?

I regret these questions have been almost completely ignored or botched in the literature on artificial intelligence. Instead, artificial intelligence (AI) work on planning, reasoning and learning has been absorbed with special cases and logical forms, and the literature on learning causal structure has, until very recently, been a poor joke. There are no villains, but there are some responsible parties. One is the hacker who cooks a program for a special case and then pretends that something general has been discovered; another is the logician turned computer scientist who wants to introduce formal semantics and prove completeness theorems when the fundamental issues in planning are inductive, not deductive; another, and in some ways the most destructive, is the psychologically connected AI practitioner who automates a procedure illustrated by some subjects in some psychological experiments, without giving thought to whether the procedure is reliable, generally feasible, or optimal.

Consider Angela's problem, although science fiction, it is no more complex than many learning problems that an android might be expected to meet and less complex than learning problems confronted every day by social scientists, economists, epidemiologists, educators, and others. The inference Angela makes is completely impossible for any combination of the celebrated AI-learning algorithms or proposals for reasoning about planning. However, it is not atypical of problems an autonomous computer might face, and—except that it is too easy—it is perfectly typical of real planning and prediction problems real humans face all the time. Angela's problem is made more difficult by the fact that experiments cannot be performed, but designing experiments is simply a special case of planning, and learning causal structure with and without experiments should fall under a general theory. What can theories of nonmonotonic reasoning, inheritance reasoning, PUPS learning, chunking concept learners, ID3-style learners, or the literature on reasoning about plans, contribute to designing a robot that functions like Angela? Logical theories about plans and conditionals and belief revision formats may contribute something, to be sure, but none of them contributes a whit to the two essential problems. Chunkers may be essential to reducing large quantities of information, but they do nothing for the central problem. Psychologists' theories of causal inference, like PUPS, just get in the way. Probabilistic classifiers and standard statistical procedures—varieties of regression, for example—would lead to exactly the end they receive in the story.

However unwittingly, two parts of the statistical literature do contribute something to Angela's problem. There is a statistical literature on the representation of causal relations by directed graphs, and on the connection between such graphical causal structures, on the one side, and probability distributions on the vertices of the graph (which do double duty as random variables) on the other. There is also statistical literature that connects prediction with counterfactual claims associated with causal dependence relations, although this literature makes no use of graphical representations. The two approaches have been unified and generalized to provide a theory of causal structure, causal inference, and prediction adequate for systems without feedback among the variables. The unification provides general, asymptotically reliable algorithms for inferring causal structure from sample data, for determining the presence of unmeasured common causes, sometimes even for inferring causal relations among unmeasured variables, and for predicting the effects of manipulations. The algorithms are asymptotically reliable, provided the probability relations and graphical structure satisfy a simple and natural axiom, they make efficient use of data, and when the actual causal graph is sparse they are computationally feasible even for very large numbers of variables. Finally, the unification provides a theory of the predictable and an algorithm that determines whether the effects on one or more variables, Y, of manipulating one or more variables, X, can be predicted from the observable relations among the variables.

In the last few years, Spirtes, Scheines, Verma, Cooper, and a few others have made breakthroughs in reliably constructing causal hypotheses from data, or from a mixture of data and background knowledge. The structures obtained with discrete variables are formally Bayesian networks, but they are also causal hypotheses. A quite general understanding has emerged concerning how such networks, so interpreted, yield predictions about the effects of direct manipulations on some of the variables. The use of experimental manipulations to provide information to construct causal hypotheses is fairly well understood. Enormous amounts of work remain to be done, and lots of fundamental things—for example, feedback and variable selection—are not yet under full theoretical control. This is the work at the cynosure of the frame problem.

REFERENCE

Spirtes, P., Scheimes, R., Glymour, C., & Meek, C. (1994). *TETRAD II*, Hillsdale, NJ: Erlbaum.

CHAPTER 4

THE FRAME PROBLEM: FREEDOM OR STABILITY? WITH PICTURES WE CAN HAVE BOTH

Lars-Erik Janlert
Department of Computing Science
University of Umea
Sweden

Let us leave aside, for the moment, the philosophical problem of how anything can possibly represent anything. In this chapter, I simply assume the basic ideas of good old-fashioned artificial intelligence: that the world can be represented by symbols that can be manipulated, that this is what reasoning is, and that the task of artificial intelligence is to design and build the required representations and symbol manipulators. It is fair to say that some progress has been made along these lines; however, many hard design problems of representational systems remain to be solved, the frame problem being one of the more infamous (McCarthy & Hayes, 1969). Half of the problem with the frame problem is that there is so much confusion and disagreement on what the problem really is. The frequent identification with a general difficulty of representing change, of dynamic representation, might be passable as a first approximation, but too vague as a definition to work by. Several distinct problems relate to the representation of change; the frame problem, properly speaking, is just one of them. Even if the various problems cannot be independently solved, there is certainly a point in being aware of the distinctions and understanding their interconnections, to be able to approach the problems in a sensible way. There would be little point in "solving" the frame problem if it meant "unsolving" some other problem.

In the first part of this chapter, I ask what we really want of a representation and attempt to give some answers, however partial, sketchy, and tentative they may be. With that as a background, the second part of the chapter makes a swift review of a number of problems associated with the representation of a changing world. In the