COGNITIVE SCIENCE **16**, 123–149 (1992)

Adversarial Problem Solving: Modeling an Opponent Using Explanatory Coherence

PAUL THAGARD

Princeton University

In adversarial problem solving (APS), one must anticipate, understand and counteract the actions of an opponent. Military strategy, business, and game playing all require an agent to construct a model of an opponent that includes the opponent's model of the agent. The cognitive mechanisms required for such modeling include deduction, analogy, inductive generalization, and the formation and evaluation of explanatory hypotheses. Explanatory coherence theory captures part of what is involved in APS, particularly in cases involving deception.

> *He who can modify his tactics in relation to his opponent, and thereby succeed in winning, may be called a heaven-born captain.* (Sun Tzu, 1983, p. 29)

Many problem-solving tasks involve other people. Often, accomplishment of a task requires coordination with others, an enterprise that might be called *cooperative* problem solving. Unfortunately, however, we also face problems that require us to take into account the actions of opponents; this is *adversarial* problem solving (APS). Both kinds of *social* problem solving have been relatively neglected by cognitive science researchers who typically investigate people's performance on nonsocial problems.

This article investigates the cognitive processes required for APS in which one must anticipate and understand the actions of an opponent. A review of several domains of APS—military strategy, business, and game playing—

This research is supported by contract MDA903-89-K-0179 from the Basic Research Office of the Army Research Institute for the Behavioral and Social Sciences. Thanks to Bruce Burns, Colin Camerer, Nicholas Findler, Keith Holyoak, Greg Nelson, and Alex Viskovatoff for useful discussions and comments, and to David Gochfield for help with the simulation in Section 4.3.

Before June 1, 1992, correspondence and requests for reprints should be sent to Paul Thagard, Cognitive Science Laboratory, Princeton University, 221 Nassau St., Princeton, NJ 08542. Thereafter, they should be sent to Department of Philosophy, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1.

will show the importance of developing a rich mental model of the adversary. I propose a set of principles of APS that summarizes the nature of the task of using a mental model of the opponent in decision making, and I describe the cognitive mechanisms required. In particular, I show how inference based on judgments of explanatory coherence can be important both for inferring the plans of an opponent and for employing deception against an opponent. Using ECHO, a connectionist model of explanatory coherence, I present simulations of two important adversarial situations: the decision by the *USS Vincennes* in 1988 to shoot down what turned out to be an Iranian airliner, and the elaborate deception employed by the Allies in World War II to convince Hitler that the Normandy invasion on D-Day was not the main Allied invasion.

1. DOMAINS OF APS

From diplomats negotiating settlements between their countries to spouses contemplating divorce, people often have to make decisions based on what they expect an actual or potential adversary to do. The most intense domain of APS is war, in which the motivation of each adversary is to dominate or destroy the other. Business competition is a less extreme kind of APS, but one that involves cognitive principles similar to those in military problem solving. Games such as chess and poker also involve anticipation of the actions of an opponent, but cognitive science and game theory have underestimated the complexity of the reasoning involved even in straightforward games. These three domains by no means exhaust the extent of APS, which can turn up in everyday life in many guises, from disputes at city hall to faculty meetings. But examples from these domains will illustrate fundamental principles of APS, particularly the importance of modeling the opponent.

1.1 Military Strategy

Over the centuries, war has consumed a lamentable portion of human resources. Hence military strategy has received much discussion, and a review of some major texts displays a large concern with APS. Of course, not all military problem solving is social: Theorists from Sun Tzu through Napoleon and Liddell Hart devoted attention to problem solving involving such aspects of the environment as terrain. Nor is all social problem solving in war adversarial, because cooperation with allies and fellow soldiers is crucial to military success. But let us look at what the leading military strategists have had to say about APS.

More than 2,500 years ago, the Chinese war lord and philosopher Sun Tzu wrote *The Art of War* which is still considered a classic of military strategy. Sun Tzu placed great emphasis on having a deep understanding of the opponent:

> If you know the enemy and know yourself, you need not fear the result of a hundred battles. If you know yourself but not the enemy, for every victory gained you will also suffer a defeat. If you know neither the enemy nor yourself, you will succumb in every battle. (Sun Tzu, 1983, p. 15)

The point of knowing enemies is to be able to check their plans (p. 15). "True excellence is to plan secretly, to move surreptitiously, to foil the enemy's intentions and balk his schemes, so that at last the day may be won without shedding a drop of blood" (p. 20). Sun Tzu advised: "Appear at points that the enemy must hasten to defend; march swiftly to places where you are not expected" (p. 25). Discovering the enemy's plans is crucial:

> Though the enemy be stronger in numbers, we may prevent him from fighting. Scheme so as to discover his plan and the likelihood of success. Rouse him, and learn the principle of his activity or inactivity. Force him to reveal himself, so as to find out his vulnerable spots. Carefully compare the opposing army with your own, so that you may know where strength is superabundant and where it is deficient. (Sun Tzu, 1983, p. 28)

Hence spies are an important part of war for the purpose of divining the enemy's plans. Conversely, great pains should be taken to conceal one's own plans from the enemy:

> All warfare is based on deception. Hence, when able to attack, we must seem unable; when using our forces, we must seem inactive; when we are near, we must make the enemy believe we are far away; when far away, we must make him believe we are near. Hold out baits to entice the enemy. Feign disorder, and crush him. (Sun Tzu, 1983, p. 11)

Sun Tzu offered advice on military matters besides dealing with an opponent, such as terrain, weather, and modes of attack, but it is clear that modeling the opponent is a crucial part of his strategic recommendations. In the following, I shall offer a cognitive analysis of what is involved in this sort of modeling.

The most influential modern military treatise is *On War* by Carl von Clausewitz, published after his death in 1831. Von Clausewitz placed less emphasis on modeling the opponent than Sun Tzu, but he wrote:

> From the character, the measures, the situation of the adversary, and the relations with which he is surrounded, each side will draw conclusions by the law of probability as to the designs of the other, and act accordingly. (von Clausewitz, 1968, p. 109)

He contended that "surprise lies at the foundation of all undertakings" (p. 269), and it is obvious that planning surprises for an opponent requires having a good enough model to be able to say what the opponent does and does not expect.

One of the leading military theorists of the 20th century, B.H. Liddell Hart, was much more psychologically oriented than von Clausewitz. In his book, *Strategy*, he concluded, on the basis of a survey of 25 centuries of war, that direct attacks on the enemy are less successful than indirect ones.

> Throughout the ages, effective results in war have rarely been attained unless the approach has had such indirectness as to ensure the opponent's unreadiness to meet it. The indirectness has usually been physical, and always psychological. (Hart, 1954, p. 25)

In discussing the failure of Britain and France to anticipate Hitler's actions that led to the second World War, he asserted:

> It is wise in war not to underrate your opponents. It is equally important to understand his methods, and how his mind works. Such understanding is the necessary foundation of a successful effort to foresee and forestall his moves. . . .A nation might profit a lot if the advisory organs of government included an "enemy department," covering all spheres of war and studying the problems of the war from the enemy's point of view—so that, in this state of detachment, it might succeed in predicting what he was likely to do next. (Hart, 1954, p. 223)

He advised generals: "Try to put yourself in the enemy's shoes and think what course it is at least probable he will foresee or forestall" (p. 348). One of the advantages of the indirect method of attack is that using distractions that suggest alternative objectives helps to keep the enemy baffled about your plans (pp. 274, 343). APS thus involves both attempting to discover the plans of the opponent and concealing your own plans.

The American Civil War General Ulysses S. Grant was a master at understanding the enemy and excelled in several kinds of knowledge that made him much more successful than other Union generals. He had an analytic knowledge of past campaigns and could discuss Napoleon's exploits in great detail. According to Keegan (1987, p. 213), "Campaign study had helped him to develop the most valuable of all his aptitudes, that of seeing into the mentality of his opponents." In particular, he came to realize that the enemy was as much afraid of him as he was of them, and thereafter never experienced trepidation in the face of the enemy.

A military analyst recently discussed the "paradoxical logic" of choice in the face of an adversary (Luttwak, 1987, p. 7). Given a choice between advancing down a wide, well-paved road or a narrow, circuitous one, a commander may well choose the bad road precisely because it is bad and therefore likely to be less well guarded by the enemy. Anticipation of the actions of an opponent can dramatically affect what appears to be the best solution to a problem.

War can involve complex mixtures of cooperative and APS. In the second World War, the leaders of Britain, the U.S., and the Soviet Union struggled both to cooperate and to compete with each other, as well to overcome their major adversary, Germany. Less intensely, scientific communities exhibit interconnected webs of cooperation and competition, as scientists attempt both to work collectively to further research and compete to advance their own careers.

1.2 Business

As in war, discovering and concealing plans is important in the competitive world of business. Writers on corporate strategy recommend analyzing competitors to determine their goals, assumptions, current strategies, and capabilities. Rogers (1987) adapted Hart's recommendation about having an "enemy department" to the business world:

> The basic quality that strategic thinkers all have in common is the ability to put themselves in the opponent's shoes—to ferret out the opponent's perceptions of the battle and their underlying assumptions, to glean their way of waging war or doing business, and to use this knowledge to do what the competition doesn't expect or what they probably won't counter with force and determination. An enemy department (it could be one employed in your company or a team of executives who do it part-time) would be given the job of "becoming" the competition, analyzing your strengths and weaknesses, preparing competitive moves against you, a list of probable reactions to your offensive moves and devising other ways of outcompeting you. (Rogers, 1987, pp. 302–303)

Rogers recommended doing an analysis of the personalities and leadership styles of the competition's leaders, and wrote that the compilation of "human-resource audits" on other firms' personnel is becoming increasingly common. Hamel and Prahalad (1989, p. 67) urged companies to develop a competitor focus at every level of their organization, with employees benchmarking their efforts against the strongest competing companies. Porter (1980) provided a widely used framework for analyzing competitors by systematically examining their goals, assumptions, capabilities, and strategy. In addition, companies must learn to interpret a competitor's market signals, which may be bluffs or warnings rather than commitments to a course of action.

Modeling other people is also extremely important in negotiations where a deal must be reached at some value between the *reservation price* of the seller—the minimum selling price—and the reservation price of the buyer: the maximum buying price. Raiffa (1982) recommended:

> *Know your adversaries.* Consider what will happen to them...if no deal is struck. Speculate about their alternatives. Examine your perceptions of their reservation price; think about the uncertainties in these perceptions (and, if it is natural to you, encode them into probabilistic assessments.) Investigate their credentials, their legitimacy, their integrity. Investigate how they have negotiated in the past. (Raiffa, 1982, p. 126–127).

Business thus abounds with APS in which it is important to get a rich model of competitors. Similarly, in legal matters, it is important for lawyers to anticipate the arguments of opposing lawyers.

1.3 Games

Because they are simpler to observe and model than war or business, games such as chess and poker have been studied by cognitive scientists more than war or business. There is now a large literature on the psychology of chess (for reviews, see, Gilhooly, 1988; Holding, 1985); topics such as search strategies and memory for patterns have been the major foci of this research, however, not the modeling of opponents. In AI models, a *minimax* strategy is usually attributed to the opponent, assuming that the opponents will always maximize their position against you and model you as doing the same. The best computer chess program, *Deep Thought,* now plays at the grand-master level, but like all other chess programs I know of, it uses the same methods against all opponents. In contrast, human chess players study the past games of their opponents to try to find weaknesses in their play. For example, Gary Kasparov, the world chess champion who played and beat *Deep Thought* twice in 1989, prepared for the match by studying the characteristics of *Deep Thought*'s games (Peterson, 1989). According to a maxim of some players: "Don't play the board, play the person." Bobby Fisher's wresting of the world championship from Boris Spassky in 1972 is often attributed to Fisher's disruptions of Spassky's psychological state. Obviously, a good chess player, human or computational, has to have an immense amount of knowledge about the game and strategies for playing it, but top-level players also seem to use models of their opponents. Although chess is a game of perfect information, because all the pieces are visible, deception is possible: It is sometimes necessary to decide whether an opponent's move against a particular piece is a genuine attack or merely a feint intended to distract the player from another kind of attack.

Whereas chess researchers have neglected opponent models, psychological and computational research on poker has paid much closer attention to players' inferences about opponents. The attention is the result of the enormous importance of bluffing in poker: Players who operate in complete compliance with the laws of probability may still end up as losers because their behavior will be too easily interpreted by opponents. Bluffing serves several purposes in addition to the obvious one of trying to win a hand when one does not have the best cards. Sometimes it pays to bluff even when you do not expect thereby to steal the hand, because your opponents will later be more likely to stay in a hand when you are betting high with a winning hand if they think you might be bluffing. Bluffing in order that you will be thought to be bluffing in the future is called "advertising." Another poker term, "sandbagging," refers to betting low even when you have a very good hand in order to draw more players into the pot. A player with poor cards may, nevertheless, stay in a hand to force a showdown that reveals an opponent's hand, thereby buying information about the opponent's pattern of play. The psychological experiments of Lopes (1976) found that performance in a simplified game of poker was enhanced by the players' understanding of probabilities and by the players' model of the opponents' bluffing tendencies.

Bluffing strategies have been incorporated into AI programs that play poker. Findler (1978) described numerous programs that play poker by evaluating human and machine opponents on the basis of statistical information kept about them. For bluffing to work well, it is important to know the style of play of one's opponents. Some consistently timid players are easily bluffed, but a bluff may be wasted on a particularly stubborn or aggressive player. A later section will discuss inferential strategies for interpreting the behavior of one's opponents.

Such strategies are beyond the scope of mathematical game theory, which does not take into account the full range of cognitive operations of participants in games. In classical game theory, each player is assumed to know the preference patterns of the other players (Luce & Raiffa, 1957, p. 5). More recent work discusses games of asymmetric information, where agents are not required to be identical or to know the characteristics of other agents (Rasmusen, 1989). Game theory can handle such cases by supposing that nature picks agents' characteristics, so that a game tree has initial branches determining what they are like. This approach works fine when there are a limited number of possible characteristics, for example, when an insurance company is interested in whether a driver is safe or unsafe, but in complex adversarial cases the game tree would be combinatorially explosive. If there are n characteristics that an opponent may or may not have, then the initial part of the game tree would need to have 2^n branches, a number that quickly becomes very large as n increases. In military decision making, and even in poker where a player can be more interested in having fun than in winning money, a major part of the task is to infer what the opponent wants, generally and in a particular case.

The cognitive narrowness of game theory is evident in the much-studied game of prisoner's dilemma, in which two suspects have to decide whether or not to confess and implicate the other. In a typical example, the suspects know that if they both confess, they each get 8 years in jail, but if neither confesses they will each get only 1 year. However, if only one of them confesses, the confessor gets a sentence of only 3 months while the other is hit with 10 years. The game is often played with iterations, so that players should be able to learn patterns of play of opponents and develop new strategies. Rarely, however, do game theorists consider what would normally be going on in real cases, where abundant inferences are made by participants concerning the traits and intentions of other players. Axelrod (1987, Axelrod &

Dion, 1988) has, however, shown by computational experiments that systems can learn better strategies for iterated prisoner's dilemma than "tit for tat," the simple strategy of defecting only after the opponent has defected. Axelrod's more successful program should still be classed as behavioral rather than cognitive because it considers only the record of past behavior of other programs with which it competes, and does not make any inferences about their strategy or inferences.

2. PRINCIPLES OF APS

Having seen an abundance of examples of APS, we can try to abstract some general principles concerning the key strategies used in interacting with an opponent. These principles are still not at the level of cognitive processes: The next section will start to discuss what kinds of representation and procedures might be required to put the principles into action. In APS, you should:

1. Construct a model of the opponent, *O*, involving *O*'s situation, past behavior, general goals, value scale, degree of competitiveness, and attitude toward risk.
2. Make sure that your model of *O* includes *O*'s model of you, because *O*'s responses to your actions will depend in part on how your actions are interpreted.
3. Use this model to infer *O*'s plans and add the inferred plans to the model.
4. Use this enhanced model to infer *O*'s likely actions and likely response to your actions.
5. Combine your model of yourself, *O*, and the environment to make a decision about the best course of action.
6. In particular, use your model of *O* to predict possible effective actions that *O* might not expect, and that, therefore, would be more effective because of the element of surprise.
7. Take steps to conceal your plans from *O* and to deceive the opponent about your plans.

Only the last two principles are special to *adversarial* problem solving. Principles 1–5 also apply to cooperative problem solving where one wants to coordinate activities with others and the goals are mutual rather than conflicting. For example, when airline pilots deal with flight emergencies, it is crucial for them to communicate plans, strategies, and explanations so that they can operate with shared mental models (Orasanu, 1990). Hence the cognitive mechanisms sketched in the next section are relevant to all social problem solving.

Obviously, the preceding analysis could be extended to urge that you should have a model of *O*'s model of your model of *O*'s model of you, and so on. Although game theorists often make strong assumptions about many layers of common knowledge possessed by adversaries, cognitive limitations make it implausible that modeling of others goes much deeper than the principles just stated.

3. COGNITIVE MECHANISMS

The purpose of the examples and principles of APS so far discussed has been to indicate the desired scope of a cognitive theory of APS. To develop that theory, we need to specify the kinds of representations that are needed to support the sorts of inferences described in the seven principles, and to specify cognitive procedures for carrying out those inferences. This section therefore outlines the structures and processes required for building and running a mental model of an opponent. Examples will be drawn from military issues arising during the second World War and from the game of poker. The discussion, however, will still be at a fairly high level of generality; the next section will present a computational implementation of one of the mechanisms described here.

My goal is to characterize the mental model that a protagonist, *P,* has of an opponent, *O.* For such a model to be useful in predicting the behavior of *O,* it cannot simply be a static description of *O,* but must be able to offer a simulation of *O*'s likely behavior. Thus, I am using the highly dynamic notion of mental model developed by Holland, Holyoak, Nisbett, and Thagard (1986; cf. Gentner & Stevens, 1983; Johnson-Laird, 1983). *P*'s cognitive apparatus must include a representation of *O* and the problem situation that both *P* and *O* find themselves in, and a set of processes that allows *P* to predict and explain what *O* will do.

3.1 Representation

P's representations must include at least the following structures.

1. Propositions.
   a. About the problem situation.
   b. About *O,* including *O*'s characteristics, goals, beliefs about the problem situation, and beliefs about *P*.
2. Rules.
   a. General rules relevant to the problem situation.
   b. More particular rules about *O*.
3. Analogs.
   a. Cases of general relevance to the problem situation.
   b. Particular cases of *O*'s past behavior.

To be more concrete, suppose *P* and *O* are playing poker together. *P*'s relevant knowledge includes particular facts about the current hand (1a), general knowledge of the rules of poker (2a), and previous cases similar to the current hand (3a). Also included are a representation of *O*'s goals (1b), presumably including winning, but such goals can come in varying degrees. Some poker players view excitement, fun, social interaction, or getting drunk at least as important as winning. *P* will also need to include in the model of *O* what *P* takes to be *O*'s beliefs about the current situation, and also, if *P* is sophisticated, what *O* thinks of *P*. *P*'s knowledge of *O* can also include generalizations about *O*'s past behavior (2b) and remembered past cases of *O*'s play or behavior in similar situations (3b). *P*'s representation of *O* should include information about where *O*'s characteristics fall on scales between aggressive and meek, knowledgeable and unsophisticated, and easy-going and cutthroat. In the background to all of this information about *P* and *O* should be fairly general principles about how beliefs, desires, and intentions lead to action (Cohen & Levesque, 1990).

3.2 Inferential Processes

First of all, *P* needs to be able to infer what *O* is likely to do, either independently or in response to possible actions of *P*. The most straightforward sort of inference for doing this is *deduction:* Given a description of the problem situation and rules that apply to the situation, deduce what the outcome of the situation will be. But *deduction* is too strong a term, because rarely does one have rules that hold universally. In poker, a rule might be:

> If Player 1 makes a high bet, and Player 2 has a weak hand, then Player 2 will fold.

There are cases, however, where the rule fails, for example, because Player 2 wants to play regardless of the cost, or wants to convince Player 1 that Player 2 is a weaker player than he or she is. So the rule just stated has to be revised to say "then Player 2 will usually fold" so that an inference from it is not strictly a deduction because the conclusion does not follow necessarily from the premises. However, the general form of the rule-based inference is the same in both the deductive and the quasi-deductive cases.

Where do the rules that enable prediction come from? Some rules can be acquired directly from statements by other people, but a cognitive system should be able to form such rules from its own observations by means of inductive generalization. One may notice, for example, that *O* tends to smile awkwardly when he or she is bluffing. Such generalizations can be made from a relatively small number of instances if one has background information about the variability of people's behavior (Holland et al., 1986, chap. 8.).

Another useful method for predicting behavior of an opponent is analogical, by remembering a similar previous case of the opponent's behavior and mapping it onto the current case to predict that *O* will do something similar. Most research on analogical problem solving has been concerned with finding a solution to a new problem by analogy with a previously solved one

(Carbonell, 1983; Gick & Holyoak, 1980, 1983), but the use of analogy in prediction is more narrow. Here the point of the analogy is not to determine what the problem solver should do directly, but to project what an opponent who is an integral part of the problem situation will do. In the Persian Gulf War of 1991, American leaders resorted to many analogies to try to foresee the actions of Iran's leader, Saddam Hussein (Holyoak & Spellman, 1991). President George Bush explicitly compared him to Adolf Hitler, whereas more moderate commentators drew parallels with former Arab leaders such as Gamal Nasser. Iraq's previous actions in its war with Iran were scrutinized for hints about how it would defend itself against the allied forces determined to drive Iraq out of Kuwait. The successful attack in Operation Desert Storm, which avoided the main Iraqi fortifications, was modeled in part on the 1863 Civil War campaign at Vicksburg, where General Grant sent his troops around the Confederate front line and attacked from the side and rear (Woodward, 1991, p. 348).

Theories and computational models of how analogs are retrieved and mapped have been offered elsewhere (Holyoak & Thagard, 1989; Thagard, Holyoak, Nelson, & Gochfeld, 1990). Although analogy is indispensable in APS, it is also potentially dangerous, as implied by the adage that the war generals always prepare for is the previous one. The Iraqis, for example, were set up for the kind of frontal assault thrown against them by Iran, but completely unprepared for the rapid flanking attack that quickly demolished their forces. Analogies can, however, be used to prevent reoccurrence of previous disasters by suggesting that plans should be modified (Hammond, 1989). In planning how to conduct the war against Iraq, American military planners had the Viet Nam analog vividly in mind, and determined to use overwhelming force as soon as possible in order to avoid involvement that started slowly and escalated painfully.

Often, *O* will perform some unexpected action and it will be crucial for *P* to explain that action. Typically, explanaton will require the formation of hypotheses about the goals and plans of *O*, and these hypotheses will feed crucially into rule-based predictions of what *O* will do. The formation of hypotheses to explain puzzling occurrences is *abductive* inference (Peirce, 1931–1958; Peng & Reggia, 1990; Thagard, 1988). Suppose, in poker, *O* makes a high bet. Several hypotheses are possible that might explain the high bet:

H1. *O* has a very good hand and expects to win the pot.
H2. *O* is bluffing and wants to drive the other players out.
H3. *O* is incompetent and is betting high without much deliberation.
H4. *O* is a thrill seeker and likes throwing money around.
H5. *O* wants to donate money to the other players.

In a normal game of poker, 1 and 2 are the hypotheses most likely to be formed, by virtue of the default rules:

R1. If a player has a very good hand and expects to win the pot, then the player bets high.
R2. If a player has a mediocre hand but thinks that the other players can be made to drop out, then the player bets high.

Abductive inference involves running such rules backward to generate hypotheses: We can explain why *O* bet high either by the hypothesis derived from R1 that *O* has a very good hand or by the hypothesis derived from R2 that *O* is bluffing. I have elsewhere surveyed the growing body of research in artificial intelligence concerned with hypothesis formation in various domains (Thagard, in press-a).

The crucial question now becomes which of the available hypotheses is the best explanation? Section 4 describes a theory of explanatory coherence that integrates multiple criteria for evaluating hypotheses: Hypotheses are to be accepted on the basis of how well they cohere with the evidence to be explained and with other hypotheses (Thagard, 1989). A good instance of the role of explanatory inference in poker is found in one of the greatest bluffs ever performed. In the 1840s, miners were returning from the gold fields in California to the east coast by a slow ship, playing draw poker on deck using the gold dust they had collected as money. One player kept four cards, drawing one. As it was dealt to him, he caught a glimpse of it just before the wind caught it and blew it overboard. The player immediately dived over the side, swam to retrieve the card, and was pulled out of the water holding the card. He then bet all of his gold dust, and all the other players folded, assuming that the last card must have given him an excellent hand: This was the best explanation available for his dramatic dive overboard. It turned out, however, that his hand consisted of four clubs and a wet diamond. The bluff was so successful because the hypothesis that he had a good hand was so much more coherent with the other players' knowledge about poker and people than was the hypothesis that the player was bluffing. In a "double bluff," an agent takes an action that the opponent is intended to explain as a bluff, even though the action in fact signals the agent's actual intention. Section 4 will discuss military examples of reasoning involving explanatory coherence.

3.3 Deception

The bluffing example shows that explanatory inferences can be crucial to one of the most interesting cognitive aspects of APS: the use of deception. In war, poker, or other adversarial situations, the protagonist, *P*, will often want to mislead *O* into having false beliefs about *P*'s intentions. To do this, *P* must have a rich model of *O*, as well as a model of *O*'s model of *P*, because *P*'s actions must be such that *O* will either interpret them in the way that *P* desires or fail to interpret tham in the way that *P* does not desire. For example, a poker player *P* who is not bluffing may start to engage in very nervous behavior such as chewing on knuckles. *P* may do this so that *O* will explain *P*'s nervous behavior by the hypothesis that *P* is bluffing, which is what *P* wants *O* to believe so that *O* will stay in the pot and *P* will win more money.

Deception is commonplace in war. In 1939, Hitler had Germany's forces mount an invasion through Holland in the hopes that the French and British would abductively infer that this was his major move (Keegan, 1989). The ruse worked, for the French rushed their forces to the north, and Hitler sent tanks crashing through the Ardennes, eventually encircling the French forces. Hitler's reasoning appears to be as follows. If I move toward Holland, then the Allies will explain this move by the hypothesis that I am attacking there. Then I can attack relatively uncontested through the Ardennes Forest which the French believe to be largely impassable. Hitler exploited the use of abductive inference by the Allies to deceive them. Section 4 contains a detailed analysis of the effective deceptions used by the Allies to conceal their intentions in the invasion of Normandy in 1944. Deception was also important in the allied victory over Iraq in the 1991 Persian Gulf War. United States Marines conspicuously prepared for an amphibious landing along the Kuwaiti coast, tying down Iraqi forces, which then became irrelevant to the actual conflict that took place after more than 300,000 U.S. troops had been secretly shifted into western Saudi Arabia far from the coast.

Deception is thus an extremely important part of APS and a fascinating cognitive process, involving as it does the protagonist's model of the opponent's model of the protagonist. This meta-model, however, does not appear to require any additional kinds of representations and inference processes besides those that are part of the basic model of the opponent, because one assumes that the opponent has roughly the same kind of cognitive apparatus as oneself.

My description of cognitive mechanisms is compatible with models of adversarial planning that have other emphases. Applegate, Elsaesser, and Sanborn (1990) described an architecture for battle management applications that takes into account both adversarial plans and plan execution and reaction. Carbonell (1981) presented a simulation of "counterplanning" situations in which an actor strives to thwart the goals and plans of a second actor. Carbonell described a large number of strategies, expressed as rules, for planning and counterplanning. Sycara (1988) provided a computational model of how adversaries' conflicting goals can be resolved using negotiation. Findler (1990) considered strategy as a decision-making mechanism that observes and evaluates its environment. In the rest of this article, I shall concentrate on complex military cases of inference based on explanatory coherence.

4. APPLICATIONS OF EXPLANATORY COHERENCE

4.1 Explanatory Coherence Theory and ECHO

The theory of explanatory coherence (TEC) has been evolving in a series of publications. Here, I give an informal statement of the current principles of TEC that suffices for understanding the cases to follow. More careful statement and detailed discussion of the principles can be found in Thagard (1989, 1991, 1992).

Principle 1: Symmetry. Explanatory coherence is a symmetric relation, unlike, say, conditional probability.

Principle 2: Explanation. (a) A hypothesis coheres with what it explains, which can either be evidence or another hypothesis; (b) hypotheses that together explain some other proposition cohere with each other; and (c) the more hypotheses it takes to explain something, the less the degree of coherence.

Principle 3: Analogy. Similar hypotheses that explain similar pieces of evidence cohere.

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

Principle 5: Contradiction. Contradictory propositions are incoherent with each other.

Principle 6: Competition. If *P* and *Q* both explain a proposition, and if *P* and *Q* are not explanatorily connected, then *P* and *Q* are incoherent with each other. (*P* and *Q* are explanatorily connected if one explains the other or if together they explain something.)

Principle 7: Acceptance. The acceptability of a proposition in a system of propositions depends on its coherence with them.

Discussion or defense of these principles here would duplicate previous writings, and I will only briefly review how they are implemented in the computer program ECHO. ECHO takes as input, statements such as

```
(EXPLAIN '(H1 H2) 'E1)
```

whose interpretation is that hypotheses H1 and H2 together explain evidence E1. ECHO represents each proposition by a network node called a *unit,* and constructs links between units in accord with TEC. Whenever two propositions cohere according to TEC, ECHO places an excitatory link (with weight greater than 0) between the units that represent them. Whenever two propositions incohere according to TEC, ECHO places an inhibitory link (with weight less than 0) between them. In accord with Principle 1, symmetry, all links are symmetric. Given the preceding input, Principle 2, explanation, requires that ECHO produce excitatory links between the units representing the following pairs of propositions: H1 and E1, H2 and E1, and H1 and H2. In accord with Principle 2(c), which provides a kind of simplicity consideration, the weights among the units given the preceding input are lower than they would be if only one hypothesis had been needed in the explanation.

Principle 3 says that analogy can also be a source of coherence, and ECHO constructs the appropriate excitatory links given input that says that propositions are analogous. To implement Principle 4, data priority, ECHO takes input specifying propositions as data and constructs an excitatory link between each of those propositions and a special evidence unit. On the basis of Principles 5 and 6, contradiction and competition, ECHO constructs inhibitory links between units representing propositions that contradict each other or compete to explain other propositions. Finally, Principle 7, acceptability, is implemented in ECHO using a simple connectionist method for updating the activation of a unit based on the units to which it is linked. Units typically start with an activation level of 0, except for the special evidence unit whose activation is always 1. Activation spreads from it to the units representing data, and from them to units representing propositions that explain data, and then to units representing higher level propositions that explain propositions that explain data, and so on. Inhibitory links between units make them suppress each other's activation. Activation of each unit is updated in parallel until the network has settled, that is, all units have achieved stable activation values; this usually takes fewer than 100 cycles. Full description of the algorithms for ECHO's operation is given in Thagard (1992).

ECHO has been used to model important cases of theory evaluation from the history of science, including Lavoisier's argument for the oxygen theory, Darwin's argument for evolution by natural selection, arguments for and against continental drift, the case of Copernicus versus Ptolemy, the case of Newton versus Descartes, and contemporary debates about why the dinosaurs became extinct (Nowak & Thagard, in press-a, in press-b; Thagard, 1989, 1991, 1992; Thagard & Nowak, 1990). ECHO has also been applied to psychological experiments on how beginning students learn science (Ranney, in press; Ranney & Thagard, 1988; Schank & Ranney, 1991) and to psychological investigations of how people perceive relationships (Miller & Read, 1991; Read & Marcus-Newhall, 1991).

4.2 The *Vincennes* Incident

Let us now look in more detail at an actual case of a decision that is naturally understood in terms of explanatory coherence. On July 3, 1988, the *USS Vincennes* was involved in a battle with Iranian gunboats in the Persian Gulf. A plane that had taken off from Iran was observed to be flying toward the *Vincennes*. On the basis of the information provided to him by his officers, Captain Rogers of the *Vincennes* concluded that the plane was an attacking Iranian F-14 and shot it down. Unfortunately, the plane turned out to be a commercial flight of Iran Air 655. Nevertheless, the official investigation (Fogarty, 1988) concluded that Rogers acted in a prudent manner. An ECHO analysis of the information available to Rogers supports that conclusion, assuming, of course, that the official report gives a complete and accurate description of that information.

Rogers's decision to fire a missile at the plane depended on his evaluation of competing hypotheses concerning the plane's nature and intentions. The hypothesis that it was a commercial flight was considered and rejected in favor of the hypotheses that the plane was an F-14 and that it was attacking. Captain Rogers recalled numerous "indicators" used in declaring the plane hostile and decided to engage it (Fogarty, 1988, p. 40). From the perspective of TEC, the F-14 hypotheses were more coherent than the alternatives for several reasons. First, they explained why the plane did not respond to verbal warnings, was not flying in a commercial air corridor, was veering toward the *Vincennes*, and was reported to be descending (this report turned out to be erroneous). Second, the commercial airline hypotheses predicted (explained) the negation of this evidence. Finally, the F-14 attack could be explained by hostile Iranian intentions for which there was ample evidence.

Appendix A contains the input given to ECHO. Note that the quoted propositions are for information only: Unlike a program that would be capable of forming the hypotheses and generating hypotheses about what explains what, ECHO does not use the content of the propositions. For ease of cross-reference, I have numbered propositions in correspondence to the list in the Fogarty (1988, p. 40) report, although a few of the pieces of evidence do not appear relevant to an assessment of explanatory coherence.

Missing from this analysis is the possible impact of analogy to previous incidents: The Fogarty report mentions the *Stark* incident that involved an Iraqi attack on an American ship in 1987. One can easily imagine Rogers's reasoning that the captain of the *Stark* should have explained the behavior of an approaching Iraqi plane in terms of its hostile intentions; analogously, Rogers should have hypothesized that the plane approaching the *Vincennes* had hostile intentions. TEC (Principle 3) and ECHO can naturally model the impact that such analogies can have.

Figure 1 displays the network that ECHO creates using the input in Table 1. Descriptions of what explains what and specification of the data lead to

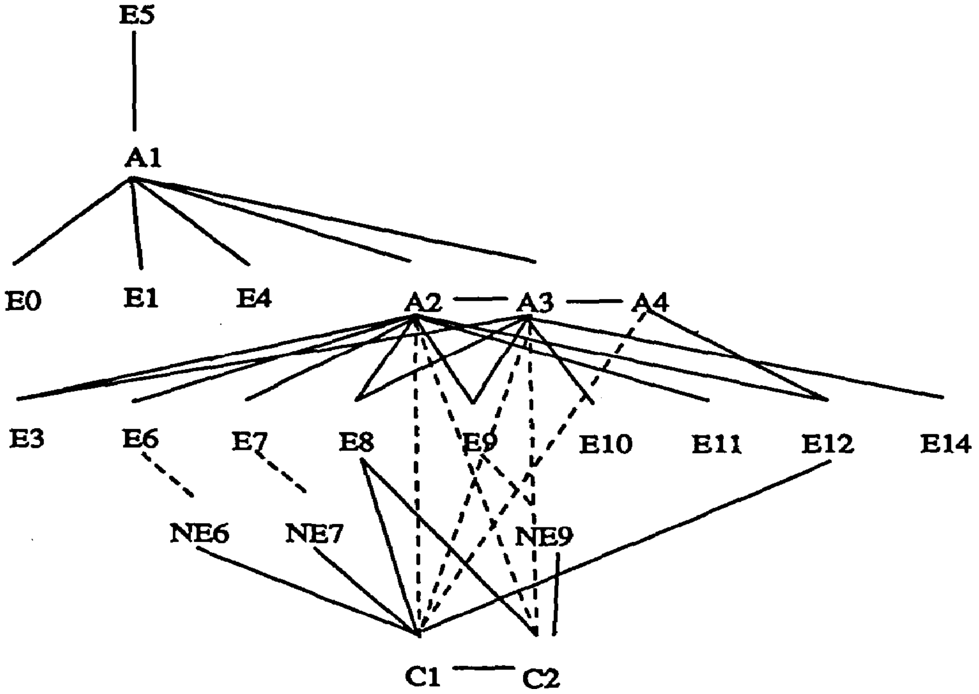

Figure 1. Network produced by ECHO in *USS Vincennes* simulation using input given in Appendix A. Straight lines indicate excitatory links produced by virtue of explanations, whereas dotted lines represent inhibitory links formed in accord with the principles of contradiction and competition. Not shown are the special evidence unit and the links with evidence units, or the link between A2 and A4 created because together they explain E12.

the creation of 45 symmetric excitatory links. ECHO also creates four symmetric inhibitory links between hypotheses that are explicitly contradictory: E6 and NE6, E7, and NE7, E9 and NE9, and A2 and C1. In addition, by virtue of the principle of competition, ECHO creates symmetric inhibitory links between four pairs of hypotheses that are not strictly contradictory, but are nevertheless competing to explain the evidence: A2 and C2, A3 and C2, A3 and C1, and A4 and C1. The "E" propositions are identified as data, so excitatory links are established between the special evidence unit and each of them. ECHO then uses a standard connectionist algorithm to update in parallel the activation of all units. After 60 cycles of updating, the activation values of all units have stabilized, and the network's decision is clear: The units representing A1, A2, A3, and A4 all have positive activation and the units representing C1 and C2 have negative activation. Thus, the propositions concerning an attacking F-14 are accepted, and propositions hypothesizing a commercial airliner are rejected, just as happened in Rogers's decision.

Because the hypothesis of an attacking F-14 was more coherent with the available information than the commercial airline hypothesis, and because F-14s were known to be capable of severely damaging the *Vincennes,* Captain Rogers shot the plane down. The fact that a tragic mistake was made does not undermine the fact that the evidence pointed strongly toward an F14. The Fogarty (1988) report found fault with the ship's Tactical Information Coordinator and Anti-Aircraft Warfare officer for providing Rogers with the erroneous information that the plane was descending rather than ascending, but this was only one factor in making the F-14 hypothesis more plausible. Thus, the decision on the *USS Vincennes* can be understood in terms of TEC, if the official report gives an accurate portrayal of the major determinants of Captain Rogers's thinking.[1]

One might think that a simpler account of Rogers's explanatory inference could be given by just saying that A2, the hypothesis that the plane is an F14, is obviously superior to C1, the hypothesis that the plane is a commercial airliner: A2 explains seven pieces of evidence but C1 only explains two. This is, of course, a large part of what accounts for Rogers's decision, but to stop there would be to neglect several components of his description of the basis of his decision. In ECHO's simulation, A2 gets activation not only from the units representing evidence, but also from the higher level hypothesis that Iran is planning an attack. The importance of having a hypothesis explained by other hypotheses is most evident in legal proceedings, which often concern questions of motivation. For example, in the U.S. Senate hearings concerning the nomination of Clarence Thomas to the Supreme Court, much attention was paid to the question of the motives of Anita Hill, who reported that Thomas had sexually harassed her. She contended that she came forward reluctantly and had no motive to do so except for a desire to express the truth. Republican senators aggressively attempted to explain her testimony by saying that she was lying or was delusional, and they suggested that she might be lying because she was part of a conspiracy against Thomas, and that she might be delusional about Thomas because of a tendency to fantasize about men. Surprisingly, the Democratic senators did not see fit to point out Thomas's strong motivation for denying her charges. Hypotheses are often accepted partly on the basis of their having higher level hypotheses that explain them, as well as on the basis of what they explain. The fact that A2 is explained by A1 is thus part of the inferential process by which A2 is accepted and the contradictory hypothesis C1 rejected.

---

[1] It is hard not to suspect that Rogers's action was also based on an assessment of the relative utility of having his ship sunk versus shooting down Iranian civilians, but the Fogarty report does not describe such as assessment.

It is also important to note that C1 is rejected not only because it contradicts A2, but also because it explains pieces of negative evidence NE6 and NE7. The form of reasoning here has the character: If the plane were a commercial airliner, it would be flying in a commercial corridor, but the plane is outside commercial corridors. The reasoning is not straightforwardly deductive, because there might be reasons why a plane is in the wrong corridor, perhaps because it is lost. In ECHO, the units representing NE6 and NE7 serve to deactivate the unit representing C1 by the following chain. E6 and E7 become active because of their links with the special evidence unit, so the activations of NE6 and NE7 drop below 0 because they have inhibitory links with E6 and E7 by virtue of the principle of contradiction. Once NE6 and NE7 have negative activation, they drag down the activation of C1 because of the symmetric excitatory links between C1 and NE6 and between C1 and NE7.

Thus, the explanatory coherence analysis of Captain Rogers's decision on the *Vincennes* shows how various factors—explanatory breadth, contradictory hypotheses, negative edvidence, and analogy—can all play a role in hypothesis evaluation. Moreover, ECHO provides a graceful way of treating these factors as soft constraints to be satisfied in parallel, resulting in an efficient means of computing which hypotheses are to be accepted and which are to be rejected. Simply counting the pieces of evidence for a hypothesis captures only part of the relevant information.[2]

4.3 Deception in the Normandy Invasion

When the Allies were planning the invasion of Normandy in 1944, they took great pains to appear to be preparing to invade at the Pas de Calais 200 miles to the east. The Allies modeled the Germans modeling the Allies: Because the Germans knew that the Allies wanted to invade the continent at a place that was close to Britain (for ease of transportation and thoroughness of air cover), the Germans were expecting the Allies to invade near the Pas de Calais. The Allies encouraged this expectation by deploying dummy landing craft and using many other tricks including a whole fictitious army group and making disinformation capturable by the Germans. Even after the D-Day invasion of Normandy, Hitler hesitated to send reinforcements to defend there because he believed that Normandy was only a feint for the real invasion that was still to come at the Pas de Calais.

---

[2] The strongest alternatives to ECHO as a computational model of complex hypothesis evaluation are based on probability theory. I have compared TEC and ECHO with the Bayesian networks of Pearl (1988) in Thagard (in press-b), and with the probabilistic approach of Peng and Reggia (1990) in Thagard (1990).

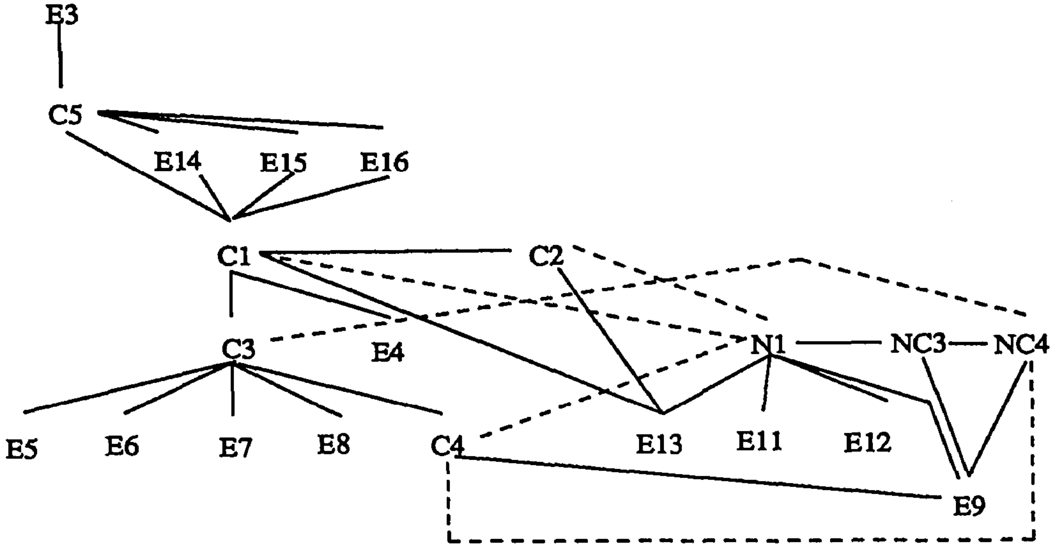

Figure 2. Network produced by ECHO in the Hitler simulation using input given in Appendix B. Straight lines indicate excitatory links produced by virtue of explanations, whereas dotted lines represent inhibitory links formed in accord with the principles of contradiction and competition. Not shown are the special evidence unit and the links with evidence units. Also omitted is the inhibitory link produced by competition between C4 and NC3, and the excitatory link between N1 and NC4 produced by their joint explanation of E9.

I now present an explanatory coherence analysis of the Allied model of the German view of the situation. Whereas in the Vincennes simulation I was modeling Captain Rogers, here I am modeling the Allies modeling Hitler. Such recursive modeling is typical of APS viewed from a cognitive perspective. Appendix B provides the input to ECHO used to model Hitler's reasoning, derived largely from Cruickshank (1979). The invasion of Normandy can be explained either by the hypothesis that it is the main Allied invasion, or by the hypotheses that it is merely a diversion and the main invasion will be at Calais. The Calais hypothesis has greater explanatory coherence than the hypothesis that the Normandy invasion is the main one for two reasons. First, that the Allies would invade at Calais is explained by its having a port, being close to England, and being closer to Germany than Normandy. Second, the Calais invasion explained numerous observations made by the Germans that were part of the Allied deception: the fake buildup in Southeast England, the simulated radio traffic, the heavy bombing of Calais, and the reports of German agents in England who, unknown to the Germans, were double agents.

ECHO uses the input from Appendix B to produce the network shown in Figure 2. By virtue of contradiction and competition, ECHO produces six inhibitory links. Pieces of evidence E3–E16 identified as data require 14 excitatory links with the special evidence unit, and an additional 25 excitatory links are created because of explanations. After 78 cycles of parallel updating of activation of each unit based on its links with other units, units representing C1–C5 become strongly activated, but N1, NC3, and NC4 are rejected. Thus, this simulation models Hitler's conclusion that the Normandy invasion was a feint.

Note the competition that occurs concerning the proper interpretation of the spies' reports. If someone says *P*, we cannot simply infer *that P*, because the person may be mistaken, lying, or joking. The truth of *P* is part of only one explanation of why *P* is said: *X* says *that P* because *X* believes *that P*, and *X* believes *that P* because *P* is true. The ECHO simulation of the German decision has them erroneously inferring that veracity is the correct explanation of the double agents' reports. As the Allies intended, the hypothesis C1 that Calais is where the main Allied invasion will take place receives support from several directions: It is explained by C5; it explains C3, E4, E13; and it coheres with C2 with which it participates in explaining E13. N1, the hypothesis that the main invasion was at Normandy, has only limited evidence for it, and contradicts important hypotheses concerning the potential Calais invasion.

4.4 Conclusion

The lesson of these simulations is that explanatory coherence considerations and ECHO can capture important parts of the inferential structure of APS. Putting together a coherent explanatory account is important whether one is attempting to infer the plans and intentions of others or planning how to make opponents make erroneous inferences about one's own intentions. TEC shows how explanatory breadth, simplicity, explanation by higher level hypotheses, competing hypotheses, analogy, and negative evidence can all affect the acceptability of a hypothesis. ECHO shows how connectionist algorithms for parallel constraint satisfaction can efficiently compute the integrated effects of the different factors affecting explanatory coherence.

Explanatory coherence considerations are only part of the cognitive operations required for APS. ECHO obviously needs to be part of a much more general cognitive system that includes rule-based and analogical reasoning, and generates the hypotheses and explanatory relations that ECHO uses to evaluate competing hypotheses.[3] But I have provided a general characterization of APS, and shown how two classes of inferences important to it, involving inferring the plans of others and modeling them to produce deception, can be understood in terms of explanatory coherence.

---

[3] For steps in this direction, see Nelson, Thagard, and Hardy (in press).

REFERENCES

Applegate, C., Elsaesser, C., & Sanborn, J. (1990). An architecture for adversarial planning. *IEEE Transactions on Systems, Man, and Cybernetics, 20,* 186–194.

Axelrod, R. (1987). The evolution of strategies in the iterated prisoner's dilemma. In L. Davis (Ed.), *Genetic algorithms and simulated annealing* (pp. 32–41). London: Pitman.

Axelrod, R., & Dion, D. (1988). The further evolution of cooperation. *Science, 242,* 1385–1390.

Carbonell, J. (1981). Counterplanning: A strategy-based model of adversary planning in real-world situations. *Artificial Intelligence, 16,* 295–329.

Carbonell, J. (1983). Learning by analogy: Formulating and generalizing plans from past experience. In R. Michalski, J. Carbonell, & T. Mitchell (Eds.), *Machine learning: An artificial intelligence approach* (pp. 137–161). Palo Alto: Tioga.

Cohen, P., & Levesque, H. (1990). Intention is choice with commitment. *Artificial Intelligence, 42,* 213–261.

Cruickshank, C. (1979). *Deception in World War* II. Oxford: Oxford University Press.

Findler, N. (1978). Computer poker. *Scientific American, 239,* 112–119.

Findler, N. (1990). *Contributions to a computer-based theory of strategies.* Berlin: Springer-Verlag.

Fogarty, W. (1988). *Formal investigation into the circumstances surrounding the downing of Iran Air Flight 655 on 3 July 1988.* Washington, DC: Department of Defense.

Gentner, D., & Stevens, A.L. (Eds.). (1983). *Mental models.* Hillsdale, NJ: Erlbaum.

Gick, M., & Holyoak, K. (1980). Analogical problem solving. *Cognitive Psychology, 12,* 306–355.

Gick, M., & Holyoak, K. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15,* 1–38.

Gilhooly, K. (1988). *Thinking* (2nd ed.). San Diego: Academic.

Hamel, G., & Prahalad, C. (1989). Strategic intent. *Harvard Business Review,* May–June, pp. 63–76.

Hammond, K. (1989). *Case-based planning: Viewing planning as a memory task.* New York: Academic.

Hart, B.H. Liddell (1954). *Strategy.* New York: Praeger.

Holding, D. (1985). *The psychology of chess.* Hillsdale, NJ: Erlbaum.

Holland, J., Holyoak, K., Nisbett, R., & Thagard, P. (1986). *Induction: Processes of inference, learning, and discovery.* Cambridge, MA: MIT Press/Bradford Books.

Holyoak, K., & Spellman, R. (1991). *If Saddam is Hitler than who is George Bush? Analogical mapping between systems of social roles.* Unpublished manuscript. Department of Psychology, University of California, Los Angeles.

Holyoak, K., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science, 13,* 295–355.

Johnson-Laird, P. (1983). *Mental models.* Cambridge, MA: Harvard University Press.

Keegan, J. (1987). *The mask of command.* New York: Viking Penguin.

Keegan, J. (1989). *The second world war.* London: Hutchinson.

Lopes, L. (1976). Model-based decision and inference in study poker. *Journal of Experimental Psychology: General, 105,* 217–239.

Luce, R.D., & Raiffa, H. (1957). *Games and decisions.* New York: Wiley.

Luttwak, E. (1987). *Strategy: The logic of war and peace.* Cambridge, MA: Harvard University Press.

Miller, L., & Read, S. (1991). On the coherence of mental models of persons and relationships. In F. FIncham & G. Fletcher (Eds.), *Cognition in close relationships* (pp. 69–99). Hillsdale, NJ: Erlbaum.

Nelson, G., Thagard, P., & Hardy, S. (in press). Integrating analogies with rules and explanations. In J. Barnden & K. Holyoak (Eds.), *Advances in Connectionism, Vol. 2: Analogical Connections*. Norwood, NJ: Ablex.

Nowak, G., & Thagard, P. (in press-a). Copernicus, Ptolemy, and explanatory coherence. In R. Giere (Ed.), *Cognitive models of science, Minnesota studies in the philosophy of science* (Vol. 15). Minneapolis: University of Minnesota Press.

Nowak, G., & Thagard, P. (in press-b). Newton, Descartes, and explanatory coherence. In R. Duschl & R. Hamilton (Eds.), *Philosophy of science, cognitive psychology, and educational theory and practice*. Albany: SUNY Press.

Orasanu, J. (1990). *Shared mental models and crew decision making* (Tech. Rep. No. 46). Princeton, NJ: Princeton University, Cognitive Science Laboratory.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. San Mateo, CA: Morgan Kaufman.

Peirce, C. (1931–1958). *Collected papers* (Vols. 1–8; edited by C. Hartshorne, P. Weiss, A. Burks). Cambridge, MA: Harvard University Press.

Peng, Y., & Reggia, J. (1990). *Abductive inference models for diagnostic problem solving*. New York: Springer Verlag.

Peterson, I. (1989). Computer chess: A masterful lesson. *Science News, 136,* 138.

Porter, M. (1980). *Competitive strategy: Techniques for analyzing industries and competitors*. New York: Free Press.

Raiffa, H. (1982). *The science and art of negotiation*. Cambridge, MA: Harvard University Press.

Ranney, M. (in press). Explorations in explanatory coherence. In E. Bar-On, B. Eylon, & Z. Schertz (Eds.), *Designing intelligent learning environments: From cognitive analysis to computer implementation*. Ablex: Norwood, NJ.

Ranney, M., & Thagard, P. (1988). Explanatory coherence and belief revision in naive physics. In *Proceedings of the 10th Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Rasmusen, E. (1989). *Games and information*. Oxford: Basil Blackwell.

Read, S., & Marcus-Newhall, A. (1991). *The role of explanatory coherence in the construction of social explanations*. Unpublished manuscript. Psychology Dept., University of Southern California.

Rogers, D. (1987). *Waging business warfare*. New York: Kensington.

Schank, P., & Ranney, M. (1991). Modeling an experimental study of explanatory coherence. In *Proceedings of the 13th Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Sun Tzu. (1983). *The art of war*. New York: Dell.

Sycara, K. (1988). Resolving goal conflicts via negotiation. In *Proceedings of the AAAI-88 Seventh National Conference on Artificial Intelligence*. San Mateo, CA: Morgan Kaufmann.

Thagard, P. (1988). *Computational philosophy of science*. Cambridge, MA: MIT Press/Bradford Books.

Thagard P. (1989). Explanatory coherence. *Behavioral and Brain Sciences, 12,* 435–467.

Thagard, P. (1990). (Review of Y. Peng and J. Reggia, *Abductive inference models for diagnostic problem solving.*) *SIGART Bulletin, 2,* (1), 72–75.

Thagard, P. (1991). The dinosaur debate: Explanatory coherence and the problem of competing hypotheses. In R. Cummins & J. Pollock (Eds.), *Philosophy and AI: Essays at the interface* (pp. 279–300). Cambridge, MA: MIT Press/Bradford Books.

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

Thagard, P. (in press-a). Hypothesis formation. In M. Keane & K. Gilhooly (Eds.), *Advances in the psychology of thinking* (Vol. 1). Hemel Hempstead: Harvester Wheatsheaf.

Thagard, P. (in press-b). Probabilistic networks and explanatory coherence. (To appear in a volume on abductive inference edited by P. O'Rourke & J. Josephson.) AAAI Press.
Thagard, P., Holyoak, K., Nelson, G., & Gochfeld, D. (1990). Analog retrieval by constraint satisfaction. *Artificial Intelligence, 46,* 259–310.
Thagard, P., & Nowak, G. (1990). The conceptual structure of the geological revolution. In J. Shrager & P. Langley (Eds.), *Computational models of discovery and theory formation.* San Mateo, CA: Morgan Kaufman.
von Clausewitz, C. (1968). *On war.* Hammondsworth: Penguin. [Original work published 1831].
Woodward, B. (1991). *The commanders.* New York: Simon & Schuster

APPENDIX A

Input to ECHO in *USS Vincennes* Simulation

Evidence

(proposition 'E0 "Gunboats were attacking the *Vincennes.*")
(proposition 'E1 "F-14s had recently been moved to Bandar Abbas.")
(proposition 'E2 "Iranian fighters had flown coincident [sic] with surface engagement on 18 April 1988.")
(proposition 'E3 "The aircraft was not responding to verbal warnings over IAD or MAD.")
(proposition 'E4 "There had been warnings of an increased threat over the July 4 weekend.")
(proposition 'E5 "There had been a recent Iraqi victory over Iran.")
(proposition 'E6 "The aircraft was not following the air corridor in the same manner as other commercial aircraft had been seen consistently to behave.")
(proposition 'NE6 "The aircraft was flying in the commercial air corridor.")
(proposition 'E7 "The aircraft was flying at a reported altitude which was lower than COMAIR was observed to fly in the past.")
(proposition 'NE7 "The aircraft flew at COMAIR's usual altitude.")
(proposition 'E8 "Track 4131 was reported to be increasing in speed.")
(proposition 'E9 "Track 4131 was reported to be decreasing in altitude.")
(proposition 'NE9 "Track 4131 was reported to be increasing in altitude.")
(proposition 'E10 "Track 4131 was CBDR to *USS Vincennes* and *USS Montgomery.*")
(proposition 'E11 "Track 4131 was reported by *USS VINCENNES'* personnel squawking Mode II-1100 which correlates with an F-14.")
(proposition 'E12 "No ESM was reflected from track 4131.")
(proposition 'E13 "F-14s have an air-to-surface capability with Maverick and modified Eagle missiles.")

```
(proposition ’E14 “The aircraft appeared to be maneuvering into attack
     position; it veered toward the USS Montgomery.”)
(proposition ’E15 “deleted in published report”)
(proposition ’E16 “Visual identification of the aircraft was not feasible.”)
(data ‘(E0 E1 E2 E3 E4 E5 E6 E7 E8 E9 E10 E11 E12 E13 E14 E15 E16))
```

Hypotheses

```
(proposition ’A1 “Iran is intending to mount an attack.”)
(proposition ’A2 “The plane is an F-14.”)
(proposition ’A3 “The plane intends to attack.”)
(proposition ’A4 “The F-14 is flying ’cold-nose’.”)
(proposition ’C1 “The plane is a commercial airliner.”)
(proposition ’C2 “The plane is taking off.”)
```

Explanations

```
(explain ’(A1) ’E0)
(explain ’(A1) ’E1)
(explain ’(A1) ’E4)
(explain ’(A1) ’A3)
(explain ’(A1) ’A2)
(explain ’(A2 A3) ’E3)
(explain ’(E5) ’A1)
(explain ’(A2) ’E6)
(explain ’(C1) ’NE6)
(explain ’(A2) ’E7)
(explain ’(C1) ’NE7)
(explain ’(A2 A3) ’E8)
(explain ’(C1 C2) ’E8)
(explain ’(A2 A3) ’E9)
(explain ’(C2) ’NE9)
(explain ’(A3) ’E10)
(explain ’(A2) ’E11)
(explain ’(A2 A4) ’E12)
(explain ’(C1) ’E12)
(explain ’(A3) ’E14)
```

Contradictions

```
(contradict ’E6 ’NE6)
(contradict ’E7 ’NE7)
(contradict ’E9 ’NE9)
(contradict ’A2 ’C1)
```

APPENDIX B

Input for Simulation of Hitler’s Reasoning About the Normandy Invasion

Evidence

```
(proposition ’E3 “The Allies want to defeat Germany quickly.”)
(proposition ’E4 “The Allies have recently bombed the region of Calais.”)
(proposition ’E5 “Activity has increased around Dover.”)
(proposition ’E6 “Landing craft are observed on the southeast coast.”)
(proposition ’E7 “Much radio traffic has been overheard in southeast
      England.”)
(proposition ’E8 “Real commanders of high stature have been assigned
      to forces in southeast England.”)
(proposition ’E9 “German agents have reported a buildup of forces in
      southeast England.”)
(proposition ’E10 “Allied diplomats say the Allies will wait until their
      forces are overwhelming.”)
(proposition ’E11 “Ships and landing craft have been observed in south-
      west England.”)
(proposition ’E12 “Allied forces are in southwest England.”)
(proposition ’E13 “The allies have invaded Normandy.”)
(proposition ’E14 “Calais is the closest point in France to England.”)
(proposition ’E15 “Calais is relatively close to Germany.”)
(proposition ’E16 “Calais has a port.”)
(data ’(E1 E2 E3 E4 E5 E6 E7 E8 E9 E10 E11 E12 E13 E14 E15 E16))
```

Hypotheses

```
(proposition ’N1 “The main invasion is at Normandy.”)
(proposition ’C1 “The main invasion will be later at Calais.”)
(proposition ’C2 “The invasion of Normandy is a diversion.”)
(proposition ’C3 “There is a large Allied force in southeast England
      preparing to invade.”)
(proposition ’NC3 “There is no large Allied force in southeast England.”)
(proposition ’C4 “The agents believe there are forces in southeast
      England.”)
(proposition ’NC4 “The agents are lying about forces in southeast
      England.”)
(proposition ’C5 “The Allies wish to reach Germany as soon as possible.”)
```

Contradictions

```
(contradict ’N1 ’C1)
(contradict ’C3 ’NC3)
(contradict ’C4 ’NC4)
```

Explanations

```
(explain '(E3) 'C5)
(explain '(C5 E15) 'C1)
(explain '(C5 E14) 'C1)
(explain '(C5 E16) 'C1)
(explain '(C1) 'C3)
(explain '(C1) 'E4)
(explain '(C3) 'E5)
(explain '(C3) 'E6)
(explain '(C3) 'E7)
(explain '(C3) 'E8)
(explain '(C3) 'C4)
(explain '(C4) 'E9)
(explain '(C1 C2) 'E13)
(explain '(N1) 'E11)
(explain '(N1) 'E12)
(explain '(N1) 'E13)
(explain '(N1 NC3 NC4) 'E9)
```