# Mental Models and Normal Errors in Naturalistic Decision Making

**Kevin Burns**

The MITRE Corporation*
202 Burlington Road
Bedford, MA 01730-1420
kburns@mitre.org

March 31, 2000
(NDM5_Burns.doc)

---

**Abstract**
This paper outlines a framework for analyzing how decision makers achieve "situation awareness" in complex, dynamic and uncertain situations. We use this framework to analyze a typical case of human "error" in the command and control of a complex system. Our analysis shows that decision errors in this and other cases can be characterized as "normal" (i.e., rational) consequences of the decision makers' mental models. We conclude by suggesting that a mental models approach can provide unifying insights into the "heuristics and biases" proposed by previous researchers.

---

## 1 Introduction

Perrow (1984) provides the following account of what he calls a "normal" accident (see Figure 1):

*"On a beautiful night in October 1978, in the Chesapeake Bay, two vessels sighted one another visually and on radar. On one of them, the Cost Guard cutter training vessel Cuyahoga, the captain (a chief warrant officer) saw the other ship up ahead as a small object on the radar, and visually he saw two lights, indicating that it was proceeding in the same direction as his own ship [Figure 1(t1), top panel]. He **thought** [emphasis added] it possibly was a fishing vessel. The first mate saw the lights, but saw three, and estimated (correctly) that it was a ship proceeding toward them [Figure 1(t1), bottom panel]. He had no responsibility to inform the captain, nor did he think he needed to. Since the two ships drew together so rapidly, the captain **decided** [emphasis added] that it must be a very slow fishing boat that he was about to overtake [Figure 1(t2), top panel]. This reinforced his incorrect interpretation. The lookout knew the captain was aware of the ship, so did not comment further as it got quite close and seemed to be nearly on a collision course. Since both ships were traveling full speed, the closing came fast. The other ship, a large cargo ship, did not establish any bridge-to-bridge communication, because the passing was routine. But at the last moment the captain of the Cuyahoga **realized** [emphasis added] that in overtaking the supposed fishing boat, which he assumed was on a near-parallel course, he would cut off that boat's ability to turn as both of them approached the Potomac River [Figure 1(t3), top panel]. So he ordered a turn to the port [Figure 1(t4), top panel]. This brought him directly in the path of the oncoming freighter, which hit the cutter [Figure 1(t4), bottom panel]. Eleven coastguardsmen perished."*

*The views and opinions expressed in this paper are those of the author, and do not necessarily reflect the views and opinions of the MITRE Corporation.

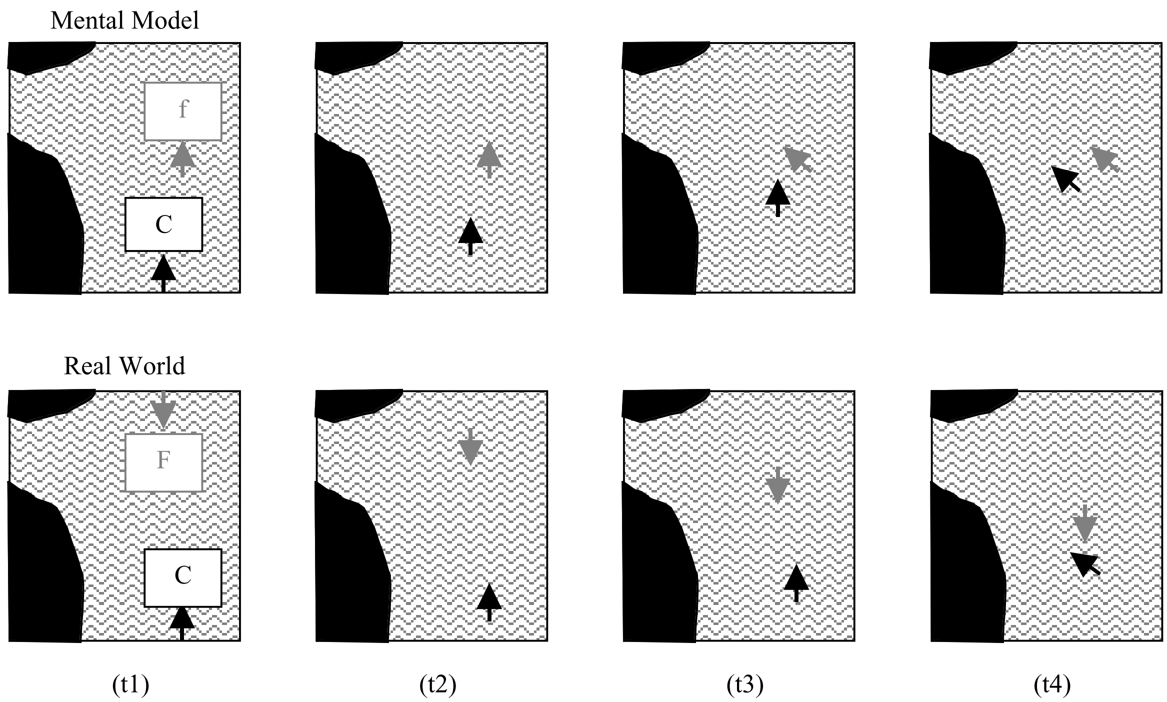


**Figure 1:** The case of the Cuyahoga. Top panels are captain's "mental model" and bottom panels are "real world". Arrow labeled "f" denotes fishing vessel (top panels) and arrow labeled letter "F" denotes Freighter (bottom panels). Arrow labeled "C" denotes Cuyahoga in top and bottom panels. Time epochs are labeled (t1), (t2), (t3) and (t4).

In retrospect (to an outside observer), the question is (Perrow 1984*): "Why would a ship in a safe passing situation suddenly turn and be impaled by a cargo ship four times its length?"* Based on his analysis of numerous accidents (in marine shipping and other complex systems, e.g., Why did operators at the Three Mile Island Nuclear Plant cut back on coolant injection and uncover the core?), Perrow concludes that the problem usually lies in the construction of a "faulty reality": *"… they [captains, operators, etc.] built perfectly reasonable* ***mental models*** *[emphasis added] of the world, which work almost all the time, but occasionally turn out to be almost an inversion of what really exists."*

This is a common theme among researchers of both human "error" (i.e., "similarity matching" and "frequency gambling", see Reason 1990) and human "power" (i.e., "intuitive recognition" and "mental simulation", see Klein 1998). In particular, Klein highlights the "perceptual" aspects of decision making (Klein and Hoffman 1993) and characterizes "recognition" as an early (and critical) step in his Recognition-Primed Decision model (Klein et al. 1989). Thus, given the importance of "situation awareness" in naturalistic decision making, and given the popular belief that situation awareness is based on "mental models", the question that motivates our research is as follows:

***Question***
What (exactly) is a mental model, and how (exactly) are mental models constructed to achieve situation awareness?

***Organization***

The remainder of this paper is organized as follows. Section 2 defines what we mean by "mental models" and "situation awareness". Sections 3 identifies three different "levels" of mental representation, and Section 4 identifies three different "stages" of mental processes that construct these representations to achieve situation awareness. Section 5 uses this framework of levels/stages to analyze the "error" committed by the captain of the Cuyahoga. Section 6 discusses why we characterize this error as "normal", and compares our approach to previous research on "heuristics and biases". Section 7 summarizes our main insights.

## 2 What is a Mental Model?

The idea that "thinking" exploits mental models was first proposed by Craik (1943, see Johnson-Laird 1983, Gregory 1987), who wrote:

*"My hypothesis then is that thought models, or parallels, reality – that its essential feature is not 'the mind', 'the self', 'sense data', nor propositions but symbolism, and that this symbolism is largely of the same kind as that which is familiar to us in mechanical devices which aid thought and calculation … the function of such symbolisation is plain. If the organism carries a 'small-scale model' of external reality and of its own possible actions within its head, it is able to try out various alternatives, conclude which is the best of them, react to future situations before they arise, utilize the knowledge of past events in dealing with the present and future, and in every way to react in a much fuller, safer, and more competent manner to the emergencies which face it."*

As expressed in these passages, Craik's theory highlights three facets of mental models that have (to a greater or lesser extent) been espoused by subsequent researchers (see Rouse and Morris 1986 for a review). First, mental models are internal belief structures (i.e., symbols) that represent an external reality. Second, mental models are constructed and manipulated by thought processes (i.e., computations) in order to accomplish cognitive tasks. Third, mental models are adapted to the context of specific situations.

Craik's theory also highlights three functions of the thought processes that operate on mental models. These three functions (noted by Rouse and Morris 1986) are to "describe", "explain" and "predict" situations. Thus, influenced by Craik's original proposal and informed by many subsequent proposals, we offer the following as a working definition of mental models:

***Definition***

*Mental models are adaptive belief constructs, used to describe, explain and predict situations.*

According to this definition, mental models are representational "structures", not computational "processes". Nevertheless, it is vacuous to talk about mental representations without also talking about how these representations are used to accomplish specific tasks. As such, our framework (Sections 3 and 4) characterizes both representational structures (which we call "mental models") and the computational processes (which we call "mental modules") that operate on these structures (see Figure 2).

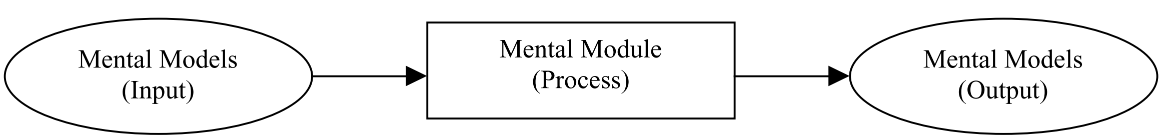


**Figure 2:** A mental module (process) operates on mental models (input) and generates mental models (output).

Also according to our definition, mental models are "constructs", which implies that they are (at one time or another) constructed. We note that some authors characterize mental models as being constructed (in a specific situation, see Wilson and Rutherford 1989), while others characterize mental models as being activated (from generalized knowledge, see Doyle and Ford 1998). Our framework (Sections 3 and 4) addresses <u>both</u> the activation (of stored models) and the construction (of new models) because we believe that both play a role in achieving situation awareness.

***What is Situation Awareness?***
We define "situation awareness" as the construction of "mental models" (see above) to describe, explain and predict a situation. This is in general agreement with the following definition of situation awareness proposed by Endsley (1988, also see Sarter and Woods 1991): *"[Situation awareness is] the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and projection of their status in the near future."*

In particular, Endsley (1988) proposes that there are three "levels" of situation awareness. We note that Endsley's three levels correspond (roughly) to the three "functions" of mental models that we identified above, i.e., "description" (of situational elements), "explanation" (of the current situation) and "prediction" (of future states and actions).

## 3 "Levels" of Mental Representation

Most theories of mental models have (to a greater or lesser extent) suggested that the mind represents information at different "levels" of abstraction. For example, Johnson-Laird (1983, working in the domain of Linguistics) proposed a hierarchy of four levels: *speech sounds* (with distinctive features), *morphemes* (which have meanings), *sentences* (which are grammatical structures) and *discourse models* (which allow inferences from meanings). Rasmussen (1985, working in the domain of Supervisory Control) proposed a hierarchy (from concrete "means" to abstract "ends") of five levels: *physical form* (nuts and bolts), *physical function* (pumps and valves), *generalised function* (pressure and temperature control), *abstract function* (mass and energy flow) and *functional purpose* (goal). Moray (1990, also working in the domain of Supervisory Control) proposed "lattice theory" as a formalism for representing four levels of "Aristotelian causation": *formal cause* (a switch causes a pump to operate because it is in the "on" position), *material cause* (because it closes a pair of contacts), *efficient cause* (because it allows current to flow to the pump) and *final cause* (because we need cooling).

These examples (also see Boer and Goodrich 1998) suggest that the number and types of "levels" used to mentally represent information will depend on the domain (and task) in which these mental models are formed (and used). In fact, generalizing the above examples, one might suggest that there are only two fundamental levels of representation (i.e., cause and effect) from which more complex hierarchies are constructed. [Here we use the term "cause" in its most general sense, to include both physical "causes" and non-physical "reasons" (see Rasmussen 1983), both of which can produce "regularities" in the world (see Richards and Bobick 1988, Richards et al. 1996).] However, human competence in uncertain situations suggests that there is another (third) level of representation that is fundamental to mental models. That is, people have at least a rudimentary (if not sophisticated) ability to represent the probability that a "cause" will produce an "effect" (see Gigerenzer et al. 1991, Johnson-Laird 1994, Burns 2000).

### ***Three Levels***
Based on the above arguments, we propose that there are three fundamental "levels" of mental models, which we characterize (from "higher" to "lower" level) as follows:

C: Confidence (which represents probability of causes and effects)
H: Hypothesis (which represents causes of effects)
E: Evidence (which represents effects)

where each level may contain a number of different "mental models" (e.g., $H_1$, $H_2$, $H_3$, etc. at the level of Hypothesis). We also propose (similar to other researchers, see above) that more complex (domain-specific) hierarchies are constructed by stacking this fundamental triad of representational "levels" into "strata" (see Figure 3). Rasmussen (1985), Moray (1990) and others have highlighted the need for human operators to work at the right stratum (which they call "level") in order to control a complex system. They have also highlighted the need for operators to move between strata in response to system perturbations. Here we stress the different levels of mental representation (i.e., C, H and E) <u>within</u> a stratum. For this reason (and to simplify our discussion), the remainder of this paper will focus on the three "levels" of representation (C, H and E) within a single "stratum".

### ***Example***
To give an example, consider the case of the Cuyahoga (see Section 1). At the start of the situation, the captain's mental model of the Evidence (E) was "two lights". One Hypothesis ($H_1$) was a "fishing vessel moving in same direction". The captain thought it was possibly a fishing vessel, where "possibly" represents some measure of Confidence in $H_1$, i.e., $C(H_1)$. In this (and many other cases), it is difficult to establish precise quantitative values for $C(H_i)$, although it is clear that people are able to articulate and act as if they represent at least rough (order-of-magnitude) probability estimates (see Burns 2000, Hertwig et al. 1999, Juslin and Olsson 1999, Richards et al. 1996, Nakayama and Shimojo 1996). Thus, for purposes of this study, we assume that $C(H_i)$ can take on the following rough (order-of-magnitude) values:

| | |
|---|---|
| 0 | Impossible |
| $\varepsilon$ | Unlikely |
| $\eta$ | Possible |
| $\rho$ | Likely |
| 1 | Certain |

For example, if the set $\{H_i\}$ includes only two Hypotheses, then $\{\varepsilon, \rho\}$ would be approximately {10%, 90%} and $\{\eta, \eta\}$ would be approximately {50%, 50%). If the set $\{H_i\}$ included one likely ($\rho$) and three unlikely ($\varepsilon$) Hypotheses, then the (normalized) values would be approximately {76%, 8%, 8%, 8%).

### ***Context***
Finally, we note that these "mental models" (E, H, C) are sensitive to the "context" set by nature (i.e., physical realities) and nurture (i.e., past experiences). For example, in the case of the Cuyahoga (see above), the captain's mental model E depends on his visual acuity, and his mental models $H_1$ and $C(H_1)$ depend on his past sailing experiences, training, etc. For simplicity of notation, we will continue to write E, H and C(H) "unconditionally", with the tacit assumption that mental models always depend on an internal "context". This "internal" context, in turn, depends on an "external" context governed by "regularities" of the real world (see Richards et al. 1992, Jepson and Richards 1993, Burns 2000).

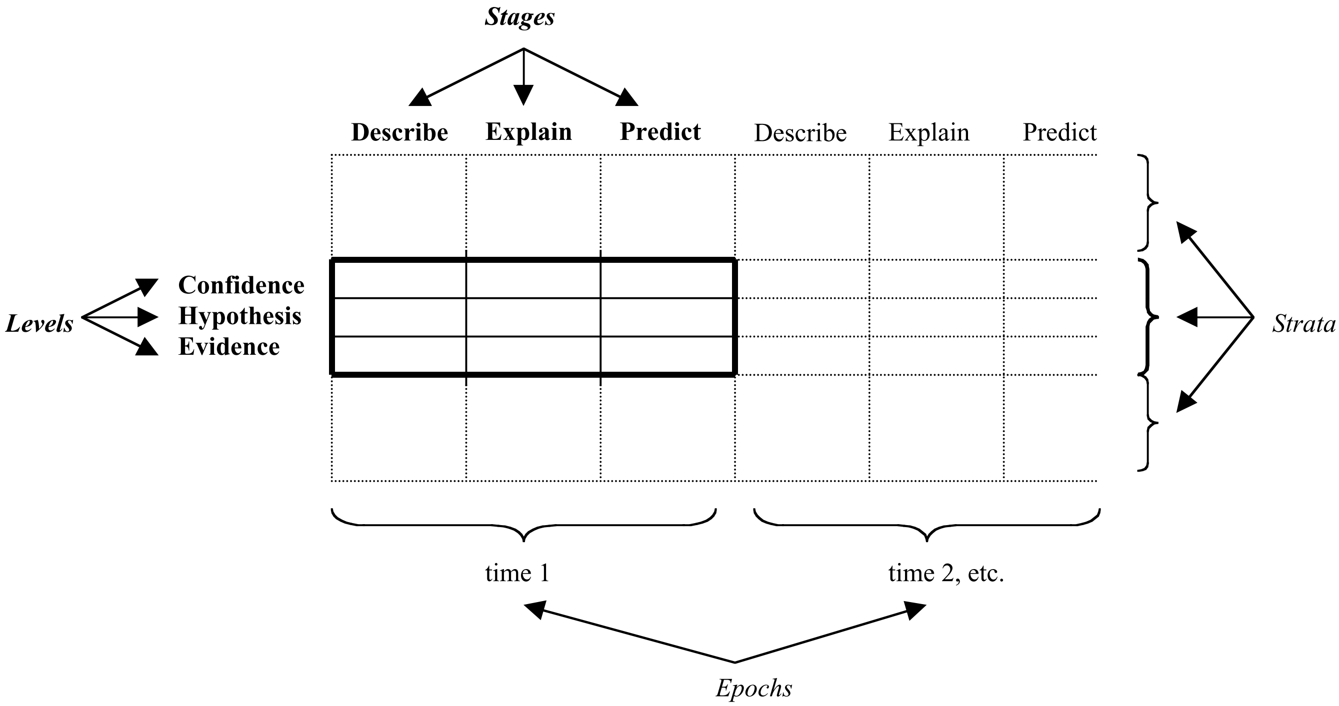


**Figure 3:** In our framework, each ***Level*** contains "mental models" (i.e., representations) and each ***Stage*** contains "mental modules" (i.e., processes). Each *Strata* contains three ***Levels*** and each *Epoch* contains three **Stages**.

## 4 "Stages" of Mental Processing

Section 2 proposed that there are three major stages of processing involved in achieving situation awareness, i.e., "describing, explaining and predicting". This section characterizes the details of each stage in terms of its component "modules" (Figure 2). Note that our discussion here is concerned with the "internal" (mental) processes of describing, explaining and predicting. The underlying assumption is that these internal processes produce internal representations (i.e., mental models) that make useful information "explicit" (see Marr 1982). These internal representations can then be "inspected" by other processes, which in turn produce external (observable) behavior, i.e., actions.

### *Describing*

The process of "describing" a situation can be characterized as "representing" information about the situation. We assume that only a limited set of Hypotheses $\{H_i\}$, i.e., those $H_i$ with high $C(H_i)$, will be activated at a given time (due to attention and memory limitations). We also assume that Confidence in each Hypothesis $C(H_i)$ is estimated (as it must be, if only $H_i$ with high $C(H_i)$ are to be activated) and represented in order-of-magnitude terms (see Section 3). Finally, we assume that Evidence is represented in a mental model E that, like $\{H_i\}$ and $\{C(H_i)\}$, is subject to resource constraints.

In summary, referring to Table 1, the "describing" stage can be characterized in terms of three component modules: "Represent Evidence", "Activate Hypotheses" and "Estimate Confidence".

***Explaining***

The process of "explaining" a situation can be characterized as identifying a Hypothesis that is "likely" to have caused the Evidence. We propose that "Bayes Rule" provides a useful formalism for analyzing this process (see Knill and Richards 1996, Burns 2000). According to Bayes Rule:

$$C(E,H_i) \;=\; C(E) * C(H_i|E) \;=\; C(H_i) * C(E|H_i)$$

where each $H_i$ is an active Hypothesis (about the cause of the Evidence). We rewrite the middle and right hand terms as:

$$C(H_i|E) \;\sim\; C(H_i) * C(E|H_i)$$

where the term C(E) has been eliminated (and the = changed to ~) because C(E) is a constant (normalizing factor) that applies to all $H_i$. The terms of this equation are commonly named as follows:

| | |
|---|---|
| $C(H_i)$ | is the "prior" (i.e., Confidence in $H_i$, before E is observed) |
| $C(E\|H_i)$ | is the "likelihood" (i.e., Confidence in observing E, given $H_i$ ) |
| $C(H_i\|E)$ | is the "posterior" (i.e., Confidence in $H_i$, given E) |

Thus, given a prior $C(H_i)$ and a likelihood $C(E|H_i)$, a decision maker can "compute" a posterior $C(H_i|E)$ using Bayes Rule. This posterior allows the decision maker to "explain" what is happening in the situation, i.e., to identify the most likely $H_i$ in light of the Evidence. Note that the term $C(E|H_i)$ is "natural" in the sense that it allows people to reason in the same direction as nature works, from causes (represented by H) to effects (represented by E). Bayes Rule then allows one to "infer" Confidence in "causes given effects" (i.e., C(H|E)), based on Confidence in "effects given causes" (i.e., C(E|H)). As such, we believe that Bayes Rule is a plausible formalism for characterizing how human decision makers "explain" situations (see Richards et al. 1996, Burns 2000).

***Mental Simulation***

The above application of Bayes Rule can "explain" Evidence only to the extent that: (1) the term $C(E|H_i)$ (as well as $C(H_i)$) can be estimated, and (2) the set of active Hypotheses $\{H_i\}$ is fixed. The first condition is important because it means that the decision maker must either "recall" or "compute" a value for $C(E|H_i)$. For familiar E and $H_i$, a value for $C(E|H_i)$ may be stored in memory (and therefore available for "recall"). In general, however, a decision maker will have to "compute" $C(E|H_i)$ by a process commonly referred to as "mental simulation" (see Kahneman and Tversky 1982, Tversky and Kahneman 1982, Kahneman and Varey 1982, Thüring and Jungermann 1986, Biel and Montgomery 1986, Jungermann and Thüring 1987, Vlek and Otten 1987, Hendrickx et al. 1989, Pennington and Hastie 1993, Klein and Hoffman 1993, Johnson-Laird 1994, Klein and Crandall 1995, Dougherty et al. 1997, Klein 1998). This process is also called "counterfactual simulation" (see Siskind 1995, 1997 for a computational implementation).

With unlimited resources, a decision maker could "simulate" numerous cases of $H_i$ (i.e., "scenarios"), and then compute $C(E|H_i)$ as the frequency of these cases that produce the actual (observed) Evidence E. However, with only limited resources, a decision maker might estimate $C(E|H_i)$ based on the "ease" of constructing one (or a few) scenarios in which $H_i$ produces the observed Evidence E. Most researchers of mental simulation (see above) have suggested that, due to practical resource constraints, human decision makers use the latter strategy. Note that this strategy can be considered "normative" to the extent that the "ease" of constructing mental scenarios (in which E is caused by $H_i$) reflects the actual probability (frequency) of scenarios in which E is caused by $H_i$ .

The second condition for using Bayes Rule (i.e., fixed $\{H_i\}$) is important because the decision maker may need to change his "frame of discernment" (i.e., the set of Hypotheses) as a situation unfolds. For example, if a Bayesian update produces posteriors $C(H_i|E)$ that are radically different from the priors (and/or if the posteriors do not adequately distinguish Confidence in different Hypotheses), then the decision maker may wish to generate a new (more discriminating) set of Hypotheses $\{H'_i\}$ and corresponding Confidences $\{C(H'_i)\}$. Thus, we can characterize two different kinds of "explanation": (1) simple explanations that involve the "updating" of Confidence in current Hypotheses, and (2) more complex explanations that involve the "generating" of new Hypotheses and associated Confidences. This latter process of generating Hypotheses (i.e., "stories", see Pennington and Hastie 1993) also involves "mental simulation" (see Klein 1998).

In summary, referring to Table 1, the "explaining" stage can be characterized in terms of three component modules: "Estimate Likelihood", "Update Confidence" and "Generate Hypotheses".

***Predicting***

The process of "predicting" a situation can be characterized as identifying what Evidence is "likely" to be observed in the future. Mental simulation (see above) plays a key role in this stage of situation awareness. That is, to predict future Evidence, a decision maker must "project" (i.e., via mental simulation, or perhaps via recall from memory) how $H_{i(t)}$ will evolve to $H_{j(t+1)}$. [Note that we use the subscript "j" because the set of Hypotheses $\{H_{j(t+1)}\}$ at time t+1 may have more (or fewer) members that the set of Hypotheses $\{H_{i(t)}\}$ at time t. For example, if each situation at time t can lead to several possible situations at time t+1, then $j > i$.] Since there may not be a one-to-one correspondence between $H_{i(t)}$ and $H_{j(t+1)}$, the decision maker must also estimate Confidence for each projected Hypothesis, i.e., $C(H_{j(t+1)})$.

Next, the decision maker must mentally simulate (or perhaps recall from memory) what Evidence $E_{k(t+1)}$ is likely to be caused by $H_{j(t+1)}$, i.e., he must estimate $C(E_{k(t+1)}|H_{j(t+1)})$. [Note that we use the subscript "k" because each Hypothesis $H_j$ may result in several cases of simulated Evidence $E_k$.] Then the decision maker can use the following version of Bayes Rule to estimate Confidence in a specific situation at time t+1 (i.e., Evidence $E_{k(t+1)}$ <u>and</u> corresponding Hypothesis $H_{j(t+1)}$):

$$C(E_{k(t+1)}, H_{j(t+1)}) = C(H_{j(t+1)}) * C(E_{k(t+1)}|H_{j(t+1)})$$

Note that the term $C(E_{k(t+1)}|H_{j(t+1)})$ that appears on the right hand side of this equation is of the same form as the "likelihood" term $C(E|H)$ discussed in the "explaining" stage (see above). Finally, the decision maker can compute his Confidence in actually observing the simulated Evidence $E_{k(t+1)}$ by summing over all active Hypotheses that might cause this Evidence:

$$C(E_{k(t+1)}) = \Sigma_j\, C(E_{k(t+1)}, H_{j(t+1)})$$

In summary, referring to Table 1, the "predicting" stage can be characterized in terms of three component modules: "Project Hypotheses", "Estimate Likelihood" and "Anticipate Evidence".

***Repeat***

The above three "stages" (i.e., describing, explaining and predicting) are repeated at each epoch (i.e., discrete time step) as new evidence is collected from an evolving situation. Thus, we characterize "situation awareness" as the dynamic construction of "mental models" via the iterative application of "mental modules" (see Table 1).

**Table 1: Mental Models and Mental Modules in Situation Awareness.**

| Stage | Module | Input Models | Output Models |
|---|---|---|---|
| Describing | | | |
| | Represent Evidence | $\{E\}$ | $E$ |
| | Activate Hypotheses | $\{H\}$ | $\{H_i\}$ |
| | Estimate Confidence | $\{C(H)\}$ | $\{C(H_i)\}$ |
| Explaining | | | |
| | Estimate Likelihood | $E, H_i$ | $C(E\|H_i)$ |
| | Update Confidence | $C(H_i), C(E\|H_i)$ | $C(H_i\|E)$ |
| | Generate Hypotheses | $H_i, C(H_i)$ | $H'_i, C(H'_i)$ |
| Predicting | | | |
| | Project Hypotheses | $H_{i(t)}, C(H_{i(t)})$ | $H_{j(t+1)}, C(H_{j(t+1)}), E_{k(t+1)}$ |
| | Estimate Likelihood | $E_{k(t+1)}, H_{j(t+1)}$ | $C(E_{k(t+1)}\|H_{j(t+1)})$ |
| | Anticipate Evidence | $C(H_{j(t+1)}), C(E_{k(t+1)}\|H_{j(t+1)})$ | $C(E_{k(t+1)})$ |

## 5 The Case of the Cuyahoga and the Freighter

Here we use our framework of mental models and mental modules (see Table 1) to analyze the case of Cuyahoga (see Section 1). The analysis is organized by time "epochs" (i.e., (t1), (t2), etc.) at which significant events are documented in Perrow's (1984) account. Within each epoch, we address each stage of mental processing (i.e., "describing, explaining and predicting") and each level of mental representation (i.e., C, H and E). We focus on the captain's mental models.

***Describing (t1)***

At (t1), the captain's mental model of the Evidence can be characterized as follows:

$E_{(t1)}$ = "two lights"

Based on Perrow's account, we assume that the captain entertained the following four Hypotheses:

$H_{1(t1)}$ = "fishing vessel moving in same direction"
$H_{2(t1)}$ = "fishing vessel moving in opposite direction"
$H_{3(t1)}$ = "other vessel moving in same direction"
$H_{4(t1)}$ = "other vessel moving in opposite direction"

Assuming that the captain considered all $H_{i(t1)}$ (roughly) equally likely at this point, we have:

$$C(H_{1(t1)}) \sim \eta$$
$$C(H_{2(t1)}) \sim \eta$$
$$C(H_{3(t1)}) \sim \eta$$
$$C(H_{4(t1)}) \sim \eta$$

where the sign ~ means that the Confidence values are order-of-magnitude and properly normalized (see Section 3). This completes our characterization of the captain's mental models E, H and C at (t1).

***Explaining (t1)***

Since "two lights" indicates "same direction" (and "three lights" indicates "opposite direction"), the observed Evidence $E_{(t1)}$ has the following "likelihood" given each Hypothesis:

$$C(E_{(t1)}|H_{1(t1)}) \sim \rho$$
$$C(E_{(t1)}|H_{2(t1)}) \sim \varepsilon$$
$$C(E_{(t1)}|H_{3(t1)}) \sim \rho$$
$$C(E_{(t1)}|H_{4(t1)}) \sim \varepsilon$$

where $\rho$ means "likely" and $\varepsilon$ means "not likely". Using Bayes Rule to update the "priors" $C(H_{i(t1)})$:

$$C(H_{1(t1)}|E_{(t1)}) \sim \eta * \rho$$
$$C(H_{2(t1)}|E_{(t1)}) \sim \eta * \varepsilon$$
$$C(H_{3(t1)}|E_{(t1)}) \sim \eta * \rho$$
$$C(H_{4(t1)}|E_{(t1)}) \sim \eta * \varepsilon$$

Thus, given the priors $C(H_{i(t1)})$ and Evidence $E_{(t1)}$, the two "most likely" (posterior) explanations at time (t1) are $H_{1(t1)}$ and $H_{3(t1)}$. Since the "posteriors" at (t1) will act as "priors" at (t2), we now set:

$$C(H_{1(t2)}) \sim \eta * \rho$$
$$C(H_{2(t2)}) \sim \eta * \varepsilon$$
$$C(H_{3(t2)}) \sim \eta * \rho$$
$$C(H_{4(t2)}) \sim \eta * \varepsilon$$

***Predicting (t1)***

At this point, the captain "mentally simulates" (or recalls from memory) what will happen in the next epoch, and constructs mental models of the resulting Hypotheses and simulated Evidence at time (t2). We summarize the simulated Evidence (for each Hypothesis) as follows:

$E_{1(t2)}$ = "ships appear to draw together rapidly"
$E_{2(t2)}$ = "ships appear to draw together rapidly"
$E_{3(t2)}$ = "ships appear to draw together slowly (or not at all)"
$E_{4(t2)}$ = "ships appear to draw together rapidly"

Note that, in this case, several Hypotheses are expected to produce the same evidence (i.e., "ships appear to draw together rapidly"). We use $E_{rapidly(t2)}$ and $E_{slowly(t2)}$ to denote the two possibilities, and we characterize the "likelihood" of observing future Evidence (given each Hypothesis) as follows:

$$C(E_{rapidly(t2)}|H_{1(t2)}) \sim \rho$$
$$C(E_{rapidly(t2)}|H_{2(t2)}) \sim \rho$$
$$C(E_{rapidly(t2)}|H_{3(t2)}) \sim \varepsilon$$

$C(E_{rapidly(t2)}|H_{4(t2)}) \sim \rho$

$C(E_{slowly(t2)}|H_{1(t2)}) \sim \varepsilon$
$C(E_{slowly(t2)}|H_{2(t2)}) \sim \varepsilon$
$C(E_{slowly(t2)}|H_{3(t2)}) \sim \rho$
$C(E_{slowly(t2)}|H_{4(t2)}) \sim \varepsilon$

Combining these "likelihoods" with the new "priors" $C(H_{i(t2)})$ from "Explaining (t1)" above, and summing over all the active Hypotheses, we get:

$$C(E_{rapidly(t2)}) \sim \Sigma_j\ (C(H_{j(t2)}) * C(E_{rapidly(t2)}|H_{j(t2)})) = \eta * [(\rho*\rho)+(\varepsilon*\rho)+(\rho*\varepsilon)+(\varepsilon*\rho)]$$
$$C(E_{slowly(t2)}) \sim \Sigma_j\ (C(H_{j(t2)}) * C(E_{slowly(t2)}|H_{j(t2)})) = \eta * [(\rho*\varepsilon)+(\varepsilon*\varepsilon)+(\rho*\rho)+(\varepsilon*\varepsilon)]$$

Since $C(E_{rapidly(t2)}) > C(E_{slowly(t2)})$, it is reasonable for the captain to "expect" the ships to appear to draw together rapidly (and to direct his "attention" to the situation at time (t2) accordingly).

***Describing (t2)***
At (t2), the new Evidence (which confirms the captain's prediction, see above) is that the ships indeed appear to draw together rapidly. That is, $E_{(t2)} = E_{rapidly(t2)}$ .

***Explaining (t2)***
Given the Evidence $E_{(t2)}$ = "ships appear to draw together rapidly", the captain can use the "likelihoods" $C(E_{rapidly(t2)}|H_{j(t2)})$ that he has already computed in "Predicting (t1)" above to update the priors $C(H_{j(t2)})$ via Bayes Rule:

$C(H_{1(t2)}|E_{(t2)}) \sim (\eta * \rho) * \rho$
$C(H_{2(t2)}|E_{(t2)}) \sim (\eta * \varepsilon) * \rho$
$C(H_{3(t2)}|E_{(t2)}) \sim (\eta * \rho) * \varepsilon$
$C(H_{4(t2)}|E_{(t2)}) \sim (\eta * \varepsilon) * \rho$

Thus, at this point, there is only one "likely" Hypothesis ($H_{1(t2)}$), i.e., that of a "fishing vessel moving in same direction". Although Perrow (1984) calls this "reinforcing the incorrect interpretation" (more formally called "Confirmation Bias", see Hutchins 1995), our analysis shows that it is actually a rational (i.e., "normative") interpretation of the situation in light of the captain's mental models (as we characterize them here).

***Predicting (t2)***
Focusing on the only "likely" Hypothesis (i.e., $H_{1(t2)}$), the captain mentally projects (i.e., simulates or recalls) a new model $H_{1(t3)}$. This new model represents a scenario in which the Cuyahoga will "cut off" the fishing vessel's ability to turn as it approaches the river (Figure 1(t3), top panel). Based on this "situation awareness", the captain decides to take action in order to prevent a collision. Notice that this decision (action) is reasonable in light of the captain's mental models at time (t2), and in light of his general knowledge about how fishing boats behave in this part of the Bay (i.e., his projected model $H_{1(t3)}$). The problem was that the "most likely" model $H_{1(t2)}$ did not match reality (in a significant way), and the unfortunate consequence was a collision (which was exactly what the captain's decision was intended to prevent!).

## 6 Discussion

The case of the Cuyahoga is a classic example of what is often called a "decision error" in the command and control of complex systems. However, contrary to this label, our analysis shows that there was really no "error" in the decision making process itself. That is, the captain's decision to take the fateful turn was reasonable in light of his "situation awareness" (i.e., mental models). Although this is the same conclusion as that reached by Perrow (1984, and others), our analysis goes further in showing that there was also no "error" in the underlying mental processes that constructed these mental models. That is, the captain's perception of the situation was consistent with the "normative" rules of Bayesian inference (given his initial model of the Evidence, i.e., "two lights"). Thus, rather than a "decision error", or even a "perception error", we suggest that the undesired outcome (in this and other cases) was the result of a "sensation error" (i.e., sensing "two lights" instead of "three lights").

### *Normal Errors*

Perrow calls such accidents "normal" because he believes that they are inevitable (although not necessarily anticipated, see Wagenaar and Groeneweg 1987) in complex human-machine systems. Our analysis suggests that such errors are also "normal" in a mathematical sense, i.e., because the underling mental computations (i.e., mental modules) and resulting mental representations (i.e., mental models) can be characterized as "normative" (i.e., rational) within "practical" (natural) resource constraints (see Simon 1990). Because this claim may appear to be odds with the vast literature on "heuristics and biases" (see Kahneman et al. 1982, also see Connolly et al. 2000), we provide additional discussion below to clarify our position.

### *Heuristics and Biases*

Laboratory experiments on human judgement and decision making are (usually) designed to isolate specific effects (behaviors). The results of these experiments are then (usually) characterized in terms of a "heuristic" process (which describes the decision maker's strategy) and a corresponding "bias" (relative to a theoretical "norm"). Numerous researchers working in this tradition have identified several dozen such biases (see Sage 1981).

Although this research approach has been very fruitful (see Piattelli-Palmarini 1994 for a popular review), there is a need (both theoretical and practical) to explain the numerous heuristics and biases in a more *unified* manner (see Gigerenzer 1996). There is also a need to more specifically characterize the *representations* on which heuristics operate. We suggest that our framework of mental models and mental modules can help meet these needs, as discussed (briefly) below.

### *Unifying Insights*

Many (most?) of the "heuristics" that have been identified by various researchers are kin to the three heuristics originally proposed by Tversky and Kahneman (1974), i.e., *Anchoring/Adjustment*, *Availability* and *Representativeness*. Here we note that all three of these heuristics are inherently captured (i.e., unified) by our framework of mental models and mental modules. For example, *Anchoring/Adjustment* is inherent in our module "Update Confidence", where a "prior" anchor is updated with a "likelihood" adjustment. Similarly, *Availability* is inherent is our module "Activate Hypotheses", where $C(H_i)$ is the formal measure of availability used to determine which $H_i$ are activated at a given time. The *Availability* heuristic is also inherent in our modules "Generate Hypotheses" and "Project Hypotheses", which consider the "ease" of constructing a causal model ($H'_i$ or $H_{j(t+1)}$) via the *Simulation* heuristic. Finally, the *Representativeness* and *Availability* heuristics together are inherent in our module "Estimate Likelihood", where the "representativeness" of a specific model (E or $H_i$) is formalized in terms of "Confidence" (also see Tenenbaum 1997).

Our framework also provides insight into the "biases" that result from these heuristics. In particular, the main insight from our case study is that effects like *Confirmation Bias* can be explained in terms of "normative" processes (which we characterize as mental modules) that operate on "situational" representations (which we characterize as mental models). We note that Gigerenzer et al. (1991) make a similar claim with regard to two other widely cited biases, i.e., *Overconfidence Effect* and the *Hard-Easy Effect*.

Our insights stem from an integrated analysis of <u>both</u> the mental representations and the mental processes involved in achieving situation awareness. However, compared to previous approaches (which have focused on heuristic processes), our approach is distinguished by its focus on contextual representations (i.e., mental models). Below we discuss why this is an important shift in focus.

***The Importance of Representation***

Researchers in the tradition of "heuristics and biases" (see Kahneman et al. 1982) <u>assume</u> that subjects have (or will) form certain internal representations (i.e., mental models) to solve the test problem they are given. These researchers then apply "normative" (e.g., Bayesian) algorithms to the assumed representations in order to derive the "rational" response. Finally, these researchers characterize the actual human response as a "bias" (i.e., "cognitive illusion", see Kahneman and Tversky 1996) relative to the rational "norm".

One problem with this approach is that the experimenter's <u>assumed</u> representations (i.e., their meta models of the subject's mental models) often do not adequately characterize the subject's tacit assumptions (see Nickerson 1996). For example, in the case of *Base Rate Neglect* (a commonly cited bias), an experimenter gives the subject a "prior" (base rate) and then poses a test problem that requires a Bayesian update of this prior. The subject's response (posterior) often exhibits a "bias" relative to what a "normative" update would produce. But, in fact, an experimenter cannot simply "plant" a prior in the subject's head. Rather, according to our framework, any information that the experimenter provides (including a base rate) is Evidence E that the subject uses to update his internal prior C(H). Thus, contrary to the classical view that subjects are not Bayesian, subjects can (in our view) be considered ultimate Bayesians because they do not ignore their internal priors.

Another problem with the "heuristics and biases" approach is that the experiments usually involve "artificial" decisions for which experimenters can define a "norm" (i.e., single correct answer). Researchers of "ecological" decision making (see Gigerenzer and Todd 1999, Chase et al. 1998, Gigerenzer 1996) note that many of the observed "biases" actually disappear when problems are posed in the "natural" context (i.e., frequencies vs. probabilities) that people experience in the real world. In other words, the "normative" solution (of an artificial problem) often requires "unnatural" mental representations, which (not surprisingly) subjects lack.

"Naturalistic" decision researchers (see Zsambok and Klein 1997) go even further in suggesting that real world decisions involve complex and dynamic tradeoffs for which there is no single "right" answer, and hence no such thing as a "norm" in the first place (Cohen 1993(a), 1993(b)). We echo these sentiments in suggesting that it is important for researchers to characterize the nature of mental representations, as well as the heuristic processes that operate on these representations. In short, if we are to explain and predict naturalistic decision making, then we must characterize how a decision maker's "mental models" are shaped by "natural context".

## 7 Conclusion

Our objective was to specify more precisely how decision makers construct mental models to achieve situation awareness. Toward this end, we outlined a Bayesian framework comprising three "levels" of mental representation and three "stages" of mental computation. We illustrated this framework with a case study in "naturalistic decision making", and gained unifying insights into the "heuristics and biases" proposed by previous researchers. Our main insight is that decision "errors" can be explained in terms of "normative" computations (which we characterize as mental mod<u>ules</u>) operating on "situational" representations (which we characterize as mental mod<u>els</u>).

## 8 References

Biel A, Montgomery H (1986). Scenarios in energy planning. In Brehmer B, Jungermann H, Lourens P, Sevón G, Eds. *New Directions in Research on Decision Making* (Elsevier).

Boer E, Goodrich M (1998). Mental models in micro worlds: Situated representations for the navigationally challenged. *IEEE International Conference on Systems, Man and Cybernetics* (San Diego, CA).

Burns K (2000). Mental models of line drawings. *Perception* (submitted).

Chase V, Hertwig R, Gigerenzer G (1998). Visions of rationality. *Trends in Cognitive Science* **2** (6) 206-214.

Cohen M (1993(a)). Three paradigms for viewing decision biases. In Klein G, Orasanu J, Calderwood R, Zsambok C, Eds. *Decision Making in Action: Models and Methods* (Norwood, NJ: Ablex).

Cohen M (1993(b)). The naturalistic basis of decision biases. In Klein G, Orasanu J, Calderwood R, Zsambok C, Eds. *Decision Making in Action: Models and Methods* (Norwood, NJ: Ablex).

Connolly T, Arkes H, Hammond K (2000). *Judgment and Decision Making: An Interdisciplinary Reader* (Cambridge, UK: Cambridge University Press).

Craik K (1943). *The Nature of Explanation* (Cambridge, UK: Cambridge University Press).

Dougherty M, Gettys C, Thomas R (1997). The role of mental simulation in judgments of likelihood. *Organizational Behavior and Human Decision Processes* **70** (2) 135-148.

Doyle J, Ford D (1998). Mental models concepts for systems dynamics research. *Systems Dynamics Review* **14** (1) 3-29.

Endsley M (1988). Design and evaluation for situation awareness enhancement. *Proceedings of the 32nd Annual Meeting of the Human Factors Society* **1** 97-101.

Gigerenzer G (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996). *Psychological Review* **103** (3) 592-596.

Gigerenzer G, Todd P (1999). *Simple Heuristics that Make Us Smart* (New York: Oxford University Press).

Gigerenzer G, Hoffrage U, Kleinbölting H (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review* **98** (4) 506-528.

Gregory R (1987). *The Oxford Companion to the Mind* (New York: Oxford University Press).

Hendrickx L, Vlek C, Oppewal H (1989). Relative importance of scenario information and frequency information in the judgment of risk. *Acta Psychologica* **72** 41-63.

Hertwig R, Hoffrage U, Martignon L (1999). Quick estimation: Letting the environment do the work. In Gigerenzer G, Todd P, Eds. *Simple Heuristics that Make Us Smart* (New York: Oxford University Press).

Hutchins E (1995). *Cognition in the Wild* (Cambridge, MA: MIT Press).

Jepson A, Richards W (1993). What makes a good feature? In Harris L, Jenkin M, Eds. *Spatial Vision in Humans and Robots* (Cambridge, MA: Cambridge University Press).

Johnson-Laird P (1994). Mental models and probabilistic thinking. *Cognition* **50** 189-209.

Johnson-Laird P (1983). *Mental Models: Towards a Cognitive Science of Language, Inference and Consciousness* (Cambridge, MA: Harvard University Press).

Jungermann H, Thüring M (1987). The use of mental models for generating scenarios. In Wright G, Ayton P, Eds. *Judgmental Forecasting* (Wiley).

Juslin P, Olsson H (1999). Computational models of subjective probability calibration. In Juslin P, Montgomery H, Eds. *Judgment and Decision Making: Neo-Brunswikian and Process-Tracing Approaches* (Mahwah, NJ: Lawrence Erlbaum).

Kahneman D, Tversky A (1996). On the reality of cognitive illusions. *Psychological Review* **103** (3) 582-591.

Kahneman D, Tversky A (1982). The simulation heuristic. In Kahneman D, Slovic P, Tversky A, Eds. *Judgment Under Uncertainty: Heuristics and Biases* (Cambridge, UK: Cambridge University Press).

Kahneman D, Varey C (1982). Propensities and counterfactuals: The loser that almost won. In Kahneman D, Slovic P, Tversky A, Eds. *Judgment Under Uncertainty: Heuristics and Biases* (Cambridge, UK: Cambridge University Press).

Kahneman D, Slovic P, Tversky A (1982). *Judgment Under Uncertainty: Heuristics and Biases* (Cambridge, UK: Cambridge University Press).

Klein G (1998). *Sources of Power: How People Make Decisions* (Cambridge, MA: MIT press).

Klein G, Crandall B (1995). The role of mental simulation in naturalistic decision making. In Hancock P, Flach J, Caird J, Vicente K, Eds. *Local Applications of the Ecological Approach to Human-Machine Systems* (Hillsdale, NJ: Lawrence Erlbaum).

Klein G, Hoffman R (1993). Seeing the invisible: Perceptual-cognitive aspects of expertise. In Rabinowitz M, Ed. *Cognitive Science Foundations of Instruction* (Hillsdale, NJ: Lawrence Erlbaum).

Klein G, Calderwood R, MacGregor D (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man and Cybernetics* **19** (3) 462-472.

Knill D, Richards W, Eds. (1996). *Perception as Bayesian Inference* (Cambridge, UK: Cambridge University Press).

Marr D (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* (New York: Freeman).

Moray N (1990). A lattice theory approach to the structure of mental models. *Philosophical Transactions of the Royal Society of London* **B 327** 577-583.

Nakayama K, Shimojo S (1996). Experiencing and perceiving visual surfaces. In Knill D, Richards W, Eds. *Perception as Bayesian Inference* (Cambridge, UK: Cambridge University Press).

Nickerson R (1996). Ambiguities and unstated assumptions in probabilistic reasoning. *Psychological Bulletin* **120** (3) 410-433.

Pennington N, Hastie R (1993). A theory of explanation-based decision making. In Klein G, Orasanu J, Calderwood R, Zsambok C, Eds. *Decision Making in Action: Models and Methods* (Norwood, NJ: Ablex).

Perrow C (1984). *Normal Accidents: Living with High-Risk Technologies* (Basic Books).

Piattelli-Palmarini M (1994). *Inevitable Illusions: How Mistakes of Reason Rule our Minds* (New York: Wiley).

Rasmussen J (1985). The role of hierarchical knowledge representation in decision making and system management. *IEEE Transactions on Systems, Man and Cybernetics* **15** (2) 234- 243.

Rasmussen J (1983). Skills, Rules and Knowledge; Signals, signs and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man and Cybernetics* **13** (3) 257-266.

Reason J (1990). *Human Error* (Cambridge, UK: Cambridge University Press).

Richards W, Jepson A, Feldman J (1996). Priors, preferences and categorical percepts. In Knill D, Richards W, Eds. *Perception as Bayesian Inference* (Cambridge, UK: Cambridge University Press).

Richards W, Feldman J, Jepson A (1992). From features to perceptual categories. In Hogg D, Boyle R, Eds. *British Machine Vision Conference* (Springer-Verlag).

Richards W, Bobick A (1988). Playing twenty questions with nature. In Pylyshyn Z, Ed. *Computational Processes in Human Vision: An Interdisciplinary Perspective* (Norwood, NJ: Ablex).

Rouse W, Morris N (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin* **100** (3) 349-363.

Sage A (1981). Behavioral and organizational considerations in the design of information systems and processes for planning and decision support. *IEEE Transactions on Systems, Man and Cybernetics* **11** (9) 640-678.

Sarter N, Woods D (1991). Situation awareness: A critical but ill-defined phenomenon. *The International Journal of Aviation Psychology* **1** (1) 45-57.

Simon H (1990). Invarients of human behavior. *Annual Review of Psychology* **41** 1-19.

Siskind J (1997). Visual event perception. In Ikeuchi K, Veloso M, Eds. *Symbolic Visual Learning*, Chapter 9 (New York: Oxford University Press).

Siskind J (1995). Grounding language in perception. *Artificial Intelligence Review* **8** 371-391.

Tenenbaum J (1997). Making sense of typicality. *Proceedings of the 19th Annual Conference of the Cognitive Science Society*.

Thüring M, Jungermann H (1986). Constructing and running mental models for inferences about the future. In Brehmer B, Jungermann H, Lourens P, Sevón G, Eds. *New Directions in Research on Decision Making* (Elsevier).

Tversky A, Kahneman D (1982). Causal schemas in judgments under uncertainty. In Kahneman D, Slovic P, Tversky A, Eds. *Judgment Under Uncertainty: Heuristics and Biases* (Cambridge UK: Cambridge University Press).

Tversky A, Kahneman D (1974). Judgement under uncertainty: Heuristics and biases. *Science* **185** 1124-1131.

Vlek C, Otten W (1987). Judgmental handling of energy scenarios: A psychological analysis and experiment. In Wright G, Ayton P, Eds. *Judgmental Forecasting* (Wiley).

Wagenaar W, Groeneweg J (1987). Accidents at sea: Multiple causes and impossible consequences. *International Journal of Man-Machine Studies* **27** 587-598.

Wilson J, Rutherford A (1989). Mental models: Theory and application in human factors. *Human Factors* **31** (6) 617-634.

Zsambok C, Klein G (1997). *Naturalistic Decision Making* (Mahwah, NJ: Lawrence Erlbaum).