

Multiple Mental Models, Automation Strategies, and Intelligent Vehicle Systems

Michael A. Goodrich
Computer Science Department
Brigham Young University
Provo, UT, USA

Erwin R. Boer
Nissan Cambridge Basic Research
Nissan Research and Development, Inc.
Cambridge, MA, USA

Abstract

An automobile driver interprets and responds to sensory input according to the context established by a mental model — an internal representation employed to encode, predict, and evaluate the consequences of perceived and intended changes to the operator's current state within the dynamic environment. Skilled driving is organized into behavioral quanta that correspond to separate mental models each with their own perceptually delineated operational domain. Switches between these behavioral quanta, referred to as either skill-based behaviors or simply skills, can be (a) mandatorily triggered by perceptual events that indicate the need for a different skill to assure acceptable performance, or (b) discretionarily triggered by a comparison of alternative skills that overlap the same perceptual domain but that can alter workload, increase safety, or decrease risk. Automation design should account for these behavioral quanta and the resulting switches among multiple mental models. The resulting design can then not only accommodate technological limitations, but, through experimentation, also account for the driver's mental models of (a) the proposed automation and (b) the dynamic vehicle-environment interaction as perceived by the driver. Depending upon the agreement between automation-generated behavior and human skill-based behavior, the designer will select an automation strategy that is safe and useful. Two automation strategies are considered: management by consent and management by exception. Management by exception is acceptable only when there is a match between the behavioral quanta of skilled human drivers and the operational limits of those driving tasks performed by the automation. By contrast, management by consent can be used when a mismatch occurs. Satisficing decision theory provides a framework of multiple interacting mental models that offers not only a description of switching of mental models with and without automation, but also guidelines for designers of vehicle automation systems.

1 Introduction

From our experimental work as well as from human factors literature, we can identify some minimum requirements of safe and useful automation systems:

1. Automation should safely and reliably operate within its intended limits, and these limits should be identifiable and interpretable by human drivers.

2. The transfer of authority between human and automation should be seamless, meaning neither the driver nor the automation should be required to work outside the limits of their operation.
3. The dynamic behavior of automation systems should be acceptable to human drivers.

The usefulness of a human-centered automation system, one in which human and automation share responsibility, is a function of not only the driver's understanding of the dynamic vehicle-environment interaction, but also the driver's understanding of the automation (that is, knowledge of both the operational domain and behavioral dynamics of the automation). In terms of mental models, a driver has both a mental model of their own skilled behavior in the world as well as a mental model of the automation. Distinctions in automation strategies can be motivated by the observation that much of human driving behavior is effected by a set of skilled behaviors. The driver perceives the world and selects an appropriate skill depending on the afforded conditions. Automation then becomes a means of efficiently performing a subset of driving skills, and an automation strategy determines not only the dynamics of how behaviors are switched between automated and manual skills, but also who is responsible for switching between these skills. These transitions are largely dictated by the automation strategy. In this paper, we consider two automation policies identified by Sarter [7]: management by consent and management by exception.

Provided that the model used by the automation accurately matches the human's mental model (assuming that the human's mental model adequately facilitates safe behavior) then automation should facilitate safe and acceptable behavior. When there is such a match between driver skills and automation behaviors, a *management by exception* automation strategy is appropriate wherein, once initiated, the automation is responsible for vehicle behavior unless the driver intervenes. By contrast, when there is a mismatch between expectations then limits of automated behaviors may not be identifiable by drivers. Under such conditions, a *management by consent* automation strategy is appropriate wherein, once initiated, the automation works for a limited time or limited scope and then discontinues unless the driver reinitiates. In this paper, we restrict attention to the automation of single tasks such as speed regulation or lane keeping, but not both. Automation strategies appropriate for multiple driving tasks are beyond the scope of this paper.

An important distinction should be made regarding how skilled behaviors are selected (where skilled behavior can be either automated or manual). Transitions between skills can be either mandatory or discretionary: a *mandatory transition* occurs when the environment changes in such a way that the current skill is inexpedient or unsafe, whereas a *discretionary transition* occurs when another skill is perceived as more expedient and/or more safe. When drivers choose to use automation (initiation), they are making a discretionary switch from their set of manual skills to a set of automated skills. By contrast, the transition from automation to manual behavior (termination) can be either mandatory or discretionary. Discretionary termination occurs when a driver determines that the skill can be better performed manually; in this paper, we present a theoretical framework for discretionary transitions but do not discuss them in detail. Mandatory termination, however, compels the driver to do something because the automation is incapable of handling the situation. In management by exception, the automation continues to operate even though it is incapable of handling the situation and the driver is required to mandatorily terminate (i.e., the driver must intervene). In management by consent, the automation self-terminates and the driver is required to take over control and employ an appropriate skill.

2 Multiple Mental Models Framework

Many aspects of cognitive decision-making have been described in terms of mental models [4]. A mental model is an internal representation employed to encode, predict, and evaluate the consequences of perceived and intended changes to the operator’s current state within the dynamic environment. We define a mental model \mathcal{M} as a triplet consisting of the perceived state of the environment Θ , a set of decisions or actions U , and a set of ordered consequences C that result from choosing $u \in U$ when $\theta \in \Theta$ obtains. According to this specification, a mental model not only encodes the relation between the input-action pair (θ, u) and the predicted consequence c , but also induces an evaluation of preferences among consequences (see Figure 1). In words, the mental model \mathcal{M} provides the context for meaningfully interpreting sensory information and generating purposeful behavior.

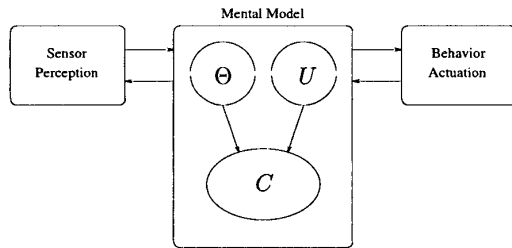


Figure 1: Working specification of a mental model.

In driving, human cognition can be described using multiple mental models (treated as agents) which can be organized into a society of interacting agents. This societal structure

not only determines which agents contribute to driver behavior, but also which agents can employ attentional resources. A three level multi-resolutional society of interacting mental models organized into a hierarchical structure (see Figure 2) can be constructed corresponding to Rasmussen’s knowledge-based (KB), rule-based (RB), and skill-based (SB) behaviors [6]. At the KB level of this hierarchy, the agent role is supervisory; at the RB level, the agent role is task management; and at the SB level, the agent role is task execution. Intuitively speaking, the KB, RB, and SB agents think, monitor, and control, respectively. These mental model agents operate within the context of overall complex human behavior.

Each mental model \mathcal{M} will be described as being enabled/disabled and engaged/disengaged. When \mathcal{M} is *enabled* the mental model is actively influencing human behavior generation, and when *disabled* the mental model has no direct influence upon behavior. When *engaged* the mental model holds attention whereby environmental information is actively perceived and interpreted, and when *disengaged* the mental model releases attention whence no such active perception occurs. In terms of Figure 1, the mental model is enabled if the arcs between the mental model and behavior/actuation are active (whence behavior u is actuated) and the mental model is engaged if the arcs between the mental model and sensor/perception are active (whence θ is actively perceived). We suppose that \mathcal{M} need not be enabled to be engaged, nor conversely.

Because automation implements skills normally performed by the driver, we discuss SB agents in more detail. The role of an SB agent is to execute a perception-based control law that performs the task specified by the RB agent. To predict and describe driver behavior, it is useful to identify computational mechanisms for coordinating a set of SB behaviors. One important aspect of this coordination is a method that identifies when a driver switches between different SB agents (i.e., how behavior is determined). For example, we have studied the conditions that trigger a switch from speed regulation to collision avoidance behaviors [1, 2] and, in the future, we will study those conditions when attention can be switched from longitudinal control to car phone usage (see Figure 2). For example (see Figure 2), in longitudi-

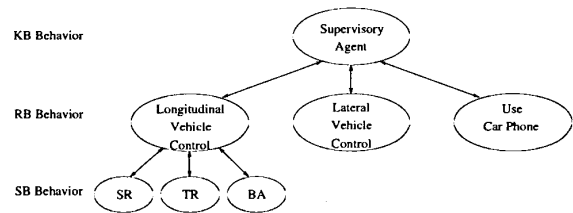


Figure 2: Hierarchical structure of agents in mental model society.

nal control there include three closed loop controllers: Speed Regulation (SR) wherein the driver regulates speed about a desired value, Time headway Regulation (TR) wherein the driver follows another vehicle at a desired time headway, and

Brake to Avoid collision (BA) wherein the driver reacts to significant dynamic disturbances such as emergency braking by a lead vehicle.

RB agents are responsible for detecting *perceptual triggering events*, operationally defined as perceived conditions mandating a switch in behavior, and evoking an appropriate response. They do so by monitoring SB agents and determining when SB behaviors are acceptable for the given environment. RB agents may also monitor the SB agents and discretionarily switch to another skill when this is appropriate. We now discuss a decision-theoretic method for identifying these mandatory and discretionary switches.

2.1 Satisficing: Mandatory Switching

Many cognitive and computer scientists recognize that insistence on optimality is a misplaced requirement in situations of limited resources and information, and that optimality inadequately describes observed behavior in naturalistic settings. Simon [8] addressed the issue of limited or bounded rationality by defining an aspiration level, such that once this level is met, the corresponding solution is deemed adequate, or *satisficing*. An important characteristic of Simon's satisficing principle is that decisions are deemed adequate on the basis of a *comparison*: any decision which exceeds the aspiration level is admissible. Satisficing Decision Theory SDT is a multi-attribute extension of this comparative characteristic which employs and compares two evaluation functions similar to the way benefit and cost are compared in economics literature. The key to this development lies in partitioning preferences over consequences into a generalized type of benefit called *accuracy* meaning *conformity to a standard*, and a generalized type of cost called *liability* meaning *susceptibility or exposure to something undesirable*.

SDT provides a method by which the accuracy and liability set membership functions can be merged: *to avoid error, a decision maker accepts those decisions which are ACCURATE and not LIABLE*. Formally, let U denote the set of possible decisions or actions and let Θ denote the states of nature. For each decision $u \in U$ and for each state of nature $\theta \in \Theta$, a consequence results which is the effect of making decision u when nature is in state θ . The accuracy $\mu_A : U \times \Theta \mapsto \mathbb{R}$ and liability $\mu_L : U \times \Theta \mapsto \mathbb{R}$ set membership functions encode the preference relations defined for each consequence (i.e., action/state-of-nature pair). The SDT decision rule may be written as

$$S_b = \{(u; \theta) : \mu_A(u; \theta) \geq b\mu_L(u; \theta)\}. \quad (1)$$

In SDT, preferences over consequences are represented by the benefit-like accuracy attribute and the cost-like liability attribute. These attributes are compared to determine when action u is admissible given state θ (i.e., when consequences are satisficing). For the speed management task, the corresponding set of driver skills includes $U = \{\text{TR}, \text{SR}, \text{BA}\}$, where TR indicates time headway regulation (car following), SR indicates speed regulation (free driving), and BA indicates active braking. Also for the speed management task,

the vector of perceptual states (see, for example, [2]) is $\theta = [T_c^{-1}, T_h, v_A]$.

Given (1), we can restrict attention to those states of nature which are satisficing for a given u , and those controls which are satisficing given the state of nature respectively defined as

$$\begin{aligned} S_b(u) &= \{\theta : \mu_A(u, \theta) \geq b\mu_L(u, \theta)\} \\ S_b(\theta) &= \{u : \mu_A(u, \theta) \geq b\mu_L(u, \theta)\}. \end{aligned}$$

In terms of behavior management by a driver, suppose a skill $u \in U$ is being used to produce behavior. The driver monitors θ via the RB mental model, and when $\theta \in S_b(u)$ no change in skill-based behavior is necessary. However, when $\theta \notin S_b(u)$, the current behavior is not acceptable and must be switched to a behavior that is appropriate for the circumstances. Given the mandate to switch, any skill $u' \in S_b(\theta)$ can be employed. An algorithm can be outlined for such rule-based task management as follows: If $\theta \in S_b(u)$ then $u' = u$; Else $u' \in S_b(\theta)$. This algorithm can be used to determine when a behavior switch is mandatory; i.e., when θ is such that u is not satisficing then a new skill $u' \neq u$ must be selected.

2.2 Domination: Discretionary Switching

For every $u \in U$ let

$$\begin{aligned} B_A(u; \theta) &= \{v \in U : \mu_L(v; \theta) < \mu_L(u; \theta) \text{ and} \\ &\quad \mu_A(v; \theta) \geq \mu_A(u; \theta)\} \\ B_L(u; \theta) &= \{v \in U : \mu_L(v; \theta) \leq \mu_L(u; \theta) \text{ and} \\ &\quad \mu_A(v; \theta) > \mu_A(u; \theta)\}, \end{aligned} \quad (2)$$

and define the set of actions that are *strictly better* than u (i.e., set of actions that dominate u)

$$B(u; \theta) = B_A(u; \theta) \cup B_L(u; \theta); \quad (3)$$

that is, $B(u; \theta)$ consists of all possible actions that have lower liability but not lower accuracy than u , or have higher accuracy but not higher liability than u . If $B(u; \theta) = \emptyset$, then no actions can be preferred to u in both accuracy and liability, and u is a (weakly) non-dominated action with respect to θ . The *non-dominated* set

$$\mathcal{E}(\theta) = \{u \in U : B(u; \theta) = \emptyset\} \quad (4)$$

contains all non-dominated actions. Elements of $\mathcal{E}(\theta)$ represent those skilled behaviors for which no other skilled behavior is obviously superior. Discretionary switches are switches from a behavior u which is satisficing to a new behavior u' which is still satisficing, but which dominates u .

The intersection of the non-dominated set with the satisficing set yields the *strongly satisficing* set

$$S_b(\theta) = \mathcal{E}(\theta) \cap S_b(\theta) \quad (5)$$

$$\mu_{S_b} = \begin{cases} \mu_{S_b} & u \in \mathcal{E}(\theta) \\ 0 & \text{otherwise} \end{cases}, \quad (6)$$

where $S_b(\theta)$ represents the crisp support set given θ . Intuitively, $S_b(\theta)$ contains only those skilled behaviors which can be done given the evidence, and \mathcal{E} does not contain any skilled behaviors which are dominated by an alternative.

2.3 Relevance to Automation

In the companion paper [3], we identified and discussed the boundaries between automated ACC behavior and human driving behavior. In that paper, we diagrammed three general cases wherein ACC behavior does not correspond to the set of driving skills. These situations are illustrated in Figures 3(a)-

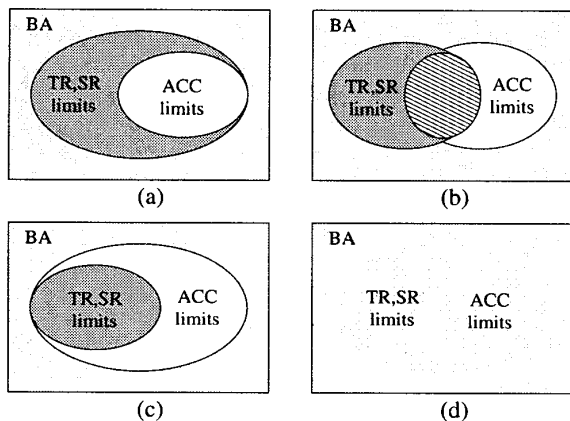


Figure 3: Comparison of TR/SR domains and ACC domain: (a) TR/SR domain broader than ACC domain $S_b(u_{ACC}) \subset S_b(TR) \cup S_b(SR)$, (b) TR/SR domain and ACC domain incompatible $S_b(u_{ACC}) \not\supseteq S_b(TR) \cup S_b(SR)$, (c) ACC domain broader than TR/SR domain $S_b(u_{ACC}) \supset S_b(TR) \cup S_b(SR)$, and (d) ACC domain approximately equals TR/SR domain $S_b(u_{ACC}) \approx S_b(TR) \cup S_b(SR)$.

(c). In these figures, the regions of support $S_b(u)$ for various skills in longitudinal control $u \in \{BA, TR, SR\}$ and the region of support (limits of automation) for an ACC system are depicted for three cases. The fourth case, which is ideal and is depicted in Figure 3(d), is $S_b(u_{ACC}) = S_b(TR) \cup S_b(SR)$ and corresponds to an ACC system that perfectly and completely automates the TR and SR longitudinal control skills.

In making the transition from automated to manual behavior, a switch is mandated when $\theta \notin S_b(u_{AUTO})$, where u_{AUTO} represents the automated behavior. (For consistency with the companion paper [3], the Figures and captions use $u_{AUTO} = u_{ACC}$ for the automation in reference to adaptive cruise control.) Since drivers are situated in real driving environments, the state θ is dynamic. The transition from $\theta \in S_b(u_{AUTO})$ to $\theta \notin S_b(u_{AUTO})$ is the *perceptual triggering event* that mandates a corresponding switch in behavior. This perceptual triggering event is naturally detected in manual driving provided¹ that the appropriate perceptual cues are receiving attention. We have elsewhere explored [3] how automation skills can use engaged manual skills to facilitate switches from automation to manual behaviors.

¹There are two factors that influence the detection of a perceptual triggering event. The first is attention to appropriate perceptual cues. The second is the ability to integrate those cues into an interpretable description of the limits of system behavior. For experienced drivers, most cognitive driving mistakes are characterized by lack of attention rather than the ability to interpret perceptual triggering events.

When one of the conditions in Figures 3(a)-(c) exists (i.e., when automation boundaries differ from transition boundaries between human skills), the driver must either learn the boundary or a surrogate assistant must facilitate detection of the perceptual triggering event. In other words, to use automation an RB mental model agent must be acquired or trained to manage switches between automated and manual skilled behavior. The detection of a perceptual triggering event presumes that the necessary perceptual cues receive attention. When a subset of skills is automated and when automation is enabled, the perceptual triggering event can be detected if the corresponding manual skills are engaged (but disabled)².

The foundational assumption behind management by exception is that the perceptual triggering event can be detected (via one of the three mechanisms: subsumption of skilled manual behaviors, learning of a new perceptual boundary, or detection via a surrogate). When this detection is not possible, an alternative must be sought. The foundational motivation behind management by consent is that a suitable alternative is to allow the automation to be invoked either for a predictable period of time or for a particular action/maneuver with the understanding that it will be shortly terminated. This induces an expectation in the operator who is aware that they are expected to “take over again in a moment” and thereby keeping the operator in the loop.

3 Case Studies: Cruise Control, ACC, and Lane Keeping

In this section, we briefly discuss three automation technologies that are used or may be used in vehicles. Based on our discussion of mental model dynamics, we discuss appropriate strategies for each of these technologies.

3.1 Cruise Control

Cruise control systems have been used in vehicles for many years. The continued installation of these systems not only attests to their usefulness in increasing driving comfort, but also demonstrates that drivers can safely detect perceptually triggering events and intervene to avoid collisions. Furthermore, we have performed experiments in which we have identified the perceptual triggering events and interpreted these events as natural transitions from the speed-regulation driver skill to either time-headway or collision-avoidance driver skills [1, 3]. We thereby conclude that attentive drivers can appropriately interpret perceptual triggering events and intervene when situations arise beyond the intended limits of automation. Furthermore, we suggest that attention to relevant perceptual cues is facilitated by the requirement that driver’s must continue to steer the vehicle. This

²If the corresponding manual skills are disengaged or if they do not exist, then either (a) the perceptual triggering event must be learned and the corresponding mental model engaged, or (b) a detectable event (perhaps via a surrogate) must be used instead. A complete discussion of this topic is beyond the scope of this paper.

prevents the driver from excessively diverting attention from the driving task. Since attention is given to relevant perceptual cues and since perceptual triggering events can be interpreted, we conclude the cruise control automation technology is sufficient to justify a management by exception automation policy.

3.2 ACC

Adaptive Cruise Control (ACC) systems are extensions to conventional cruise control that not only regulate speed about a preset value, but also control speed such that, in the presence of other traffic, time headway is regulated about a preset value. The extension of the effectiveness of cruise control as well as experimental results both suggest that adverse ACC effects cannot be explained by decreased alertness [5]; it appears that drivers are able to attend to relevant perceptual cues presumably because they must still steer the vehicle.

However, the detection of the limits of ACC capabilities appears to be a more difficult task for drivers [5]. One possible explanation for this difficulty is a mismatch between manual driving skills and limits of ACC. In an effort to increase the useful set of conditions for which ACC can be used, many proposed ACC systems include a limited braking strategy. In these limited braking strategies, the brakes are activated but maximum deceleration is limited so that the automation may brake but not decelerate more than, for example, 0.4G. There is evidence that such partial braking can adversely affect safety [5]. In the context of our multiple mental model discussion, this effect can be explained by the distinction between the active braking (BA) skill and the time-headway regulation (TR) skill. We suggest that time-headway regulation is a skill that is actuated exclusively by the accelerator pedal to accelerate or decelerate using engine braking. By contrast, whenever the brake pedal is pressed only when the BA skill is enabled. As a baseline, ACC systems perform the SR and TR skills. Some ACC systems also perform partial braking and thereby not only perform the SR and TR skills, but also perform a portion of (but not all of) the BA skill (see Figure 3(c)). Consequently, it is difficult for drivers to detect the appropriate perceptual triggering event that identifies the limits of ACC behavior. Instead, some drivers adopt a "wait and see" strategy. Two solutions to this problem are to limit ACC behavior to the SR and TR skills, or to include a warning system that acts as a surrogate perceptual triggering event detector. Regardless of which approach is taken (assuming an appropriately designed surrogate), technology appears to be sufficient to closely match the set of driver skills associated with car-following. Thus, a management by exception automation strategy is appropriate.

3.3 Lane Keeping

Emerging from the development of fully autonomous vehicles and advanced highway systems has been technology for autonomous lane keeping. It appears that some vehicle manufacturers are exploring how this technology can be applied

to augmenting a driver's ability to steer a vehicle. These approaches would likely be paired with ACC systems to create semi-autonomous vehicles. To implement a lane-keeping system without changing highway infrastructure suggests the use of an in-vehicle vision system that determines lane position. Technologically, vision systems exist that can effectively maintain lane position (solutions include adaptive region growing, lane detection, and other pattern recognition algorithms). However, the mental models in lane keeping are not as clearly delineated as they are for longitudinal control. Consequently, there will likely be mismatches between a driver's manual skills and the limits of lane-keeping behavior. This implies the need for a surrogate assistant.

Additionally, when drivers are not required to regulate speed or maintain lane position it seems unlikely that attention will be given to relevant perceptual cues. Consequently, even if a suitable surrogate is developed that allows drivers to detect perceptual triggering events, an additional system will be required to help drivers attend to appropriate perceptual cues. Until the difficulty of simultaneously satisfying these two requirements is resolved, we suggest that lane keeping should adopt a management by consent automation policy. Such a policy allows a driver to let the automation "take the steering wheel for a moment" with an understanding that the driver will shortly be required to again steer the vehicle. This understanding induces an expectation in the driver that they will be required to control the vehicle at a predictable moment in the future. Thus, a predictable and triggering event is induced not by interpreting relevant perceptual cues, but instead by an understanding of the limited scope of operation.

4 Conclusion

In this paper, we have developed a motivation for selecting an automation strategy based on an understanding of a driver's mental model dynamics. This development is based on the observation that mental models not only dictate selection of skilled driver behaviors, but also determine the initiation and termination of automation. Furthermore, this observation helps establish guidelines for the system design by identifying natural skill domains, relevant perceptual cues and attentional demands, which leads to an automation patterned after skilled driver behavior.

Automation strategies include *automation by consent* and *automation by exception* which are characterized by the termination of automation. For automation by consent, termination is initiated by the automation since the system only operates for a limited time or until one specific task is completed (e.g., changing a lane, or centering the vehicle in the center of the lane). By contrast, for automation by exception, termination is initiated by the driver when the driver decides to resume manual control (e.g., when the operational limits of automation are reached the driver intervenes). Determining which automation strategy is appropriate includes (a) defining the limits of automation and then determining if the driver can interpret the corresponding perceptual triggering event and (b) identifying if attention can and will be placed

propriate perceptual cues. If drivers attend to relevant perceptual cues and can detect the perceptual triggering event (perhaps through a surrogate) then automation by exception is an appropriate automation policy. If, however, drivers cannot attend to relevant perceptual cues or cannot detect a perceptual triggering event then automation by consent is an appropriate automation policy. This perspective should help designers of advanced vehicle systems to produce safe systems, particularly when automation and human share responsibility.

References

- [1] M. A. Goodrich and E. R. Boer. Semiotics and mental models: Modeling automobile driver behavior. In *Joint Conference on Science and Technology of Intelligent Systems ISIC/CIRA/ISAS'98 Proceedings*, Gaithersburg, MD, September 1998.
- [2] M. A. Goodrich, E. R. Boer, and H. Inoue. Brake initiation and braking dynamics: A human-centered study of desired ACC characteristics. Technical Report TR-98-5, Cambridge Basic Research, Nissan Research and Development, Inc., Cambridge, MA 02142, USA, 1998.
- [3] M. A. Goodrich, E. R. Boer, and H. Inoue. A model of human brake initiation behavior with implications for ACC design. In *ITSC 99*, 1999. To appear.
- [4] P. N. Johnson-Laird. *The Computer and the Mind: An Introduction to Cognitive Science*. Harvard University Press, Cambridge, Massachusetts, 1988.
- [5] L. Nilsson. Safety effects of adaptive cruise controls in critical traffic situations. In *Proceedings of the Steps Forward, Volume III*, pages 1254–1259, Yokohama, Japan, November 9-11, 1995. the Second World Congress on Intelligent Transport Systems.
- [6] J. Rasmussen. Outlines of a hybrid model of the process plant operator. In T. B. Sheridan and G. Johansen, editors, *Monitoring Behavior and Supervisory Control*, pages 371–383. Plenum, 1976.
- [7] N. Sarter. Making coordination effortless and invisible: The exploration of automation management strategies and implementations. Presented at the 1998 CBR Workshop on Human Interaction with Automated Systems, June 1 1998.
- [8] I. A. Simon. A behavioral model of rational choice. *Quart. J. Economics*, 59:99–118, 1955.