Model-Based Human-Centered Task Automation: A Case Study in ACC System Design

Michael A. Goodrich and Erwin R. Boer

***Abstract*—Engineers, business managers, and governments are increasingly aware of the importance and difficulty of integrating technology and humans. The presence of technology can enhance human comfort, efficiency, and safety, but the absence of human-factors analysis can lead to uncomfortable, inefficient, and unsafe systems. Systematic human-centered design requires a basic understanding of how humans generate and manage tasks. A very useful model of human behavior generation can be obtained by recognizing the task-specific role of mental models in not only guiding execution of skills but also managing initiation and termination of these skills. By identifying the human operator's mental models and using them as templates for automating different tasks, we experimentally support the hypothesis that natural and safe interaction between human operator and automation is facilitated by this model-based human-centered approach. The design of adaptive cruise control (ACC) systems is used as a case study in the design of model-based task automation systems. Such designs include identifying ecologically appropriate perceptual states, identifying perceptual triggering events for managing transitions between skilled behaviors, and coordinating the actions of automation and operator.**

***Index Terms*—Human-centered automation, mental models, satisficing.**

I. Introduction

With the rapid progress of technology, there is a concerted effort to use automated systems to augment human abilities in safety critical tasks such as automobile driving, aviation, and process automation. However, a lesson learned from process automation is that, in the absence of human factors considerations, even technologically state-of-the-art systems can be more problematic than beneficial [1], [27], [32]. This lesson indicates the importance of including human factors in the automation design process so as to prevent "ironies of automation" [1].

Human-centered automation is a magnificent ideal, but one that is difficult to achieve in practice. One reason for this difficulty is that much of human-factors analysis produces case-specific phenomenological data, that is, data reporting trends and correlations but without plausible explanations of cause and effect. This compels the individual designer to make subjective predictions and conjectures and this, in turn, causes human-centered design to often be more of an art than a science.

As an alternative to case-specific, phenomenology-based design, an integrated model of human behavior-generation allows a designer to expand attention to include not only how human operators behave, but also why they do so. Such a model must not only reflect human processing, perception, and behavioral characteristics, but also facilitate accurate predictions and descriptions of human behavior (particularly the potential behavioral modifications triggered by the introduction of automation) [3], [5]. With the availability of an integrated model, the designer is in a position to construct theory-guided experiments and model-guided designs.

From our experimental work, as well as from human factors literature, we can identify some important factors influencing the perceived safety and usefulness of automation systems. In particular, the following four factors influence how humans use automation design to support and/or take over certain aspects of skilled human behavior.

1) **Limitations**: Automation should safely and reliably operate within its intended limits, and these limits should be easily identifiable and interpretable by human operators.[1]
2) **Responsibility**: The transfer of authority between human and automation should be seamless, meaning neither the operator nor the automation should be required to work outside the limits of their operation. Additionally, the operator should know who does what work and when.
3) **Dynamics**: The dynamic execution behavior of automation systems should be acceptable[2] to and predictable by human operators under all conditions.[3]
4) **Efficiency**: The automation should perform skilled execution more efficiently than the operator and thereby relieve the operator of some of the physical and mental burden of performing the task. Additionally, operation of the system should impose minimal workload.

When these human-centered design principles are practiced, the human operator is more likely to effectively interact with the automation.

This paper is written from a perspective that *task automation systems*, defined in the next section, are intended to safely

Manuscript received October 29, 1999; revised April 23, 2003. This paper was recommended by Associate Editor A. Ollero.

M. A. Goodrich is with the Computer Science Department, Brigham Young University, Provo, UT, USA.

Erwin R. Boer is with Consulting, San Diego, CA, 92037 USA.

This work was researched while the authors were with Nissan Cambridge Basic Research, Cambridge, MA, 02139 USA.

Digital Object Identifier 10.1109/TSMCA.2003.817040

[1]This is associated with the notion of *transparency* in the literature. However, transparency seems to capture the notion that humans should be able to use automation without expending a lot of cognitive effort. In this context, each of the four factors presented herein contribute to the transparency of the automation.

[2]Acceptance by operators implies that the automation performs the task sufficiently well and that the operator can control the automation.

[3]Transitions associated with dynamic task allocation should be acceptable and predictable by operators. We choose to include elements of dynamic task allocation under *limitations* rather than *dynamics*.

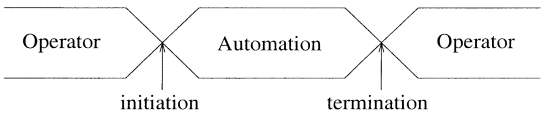


Fig. 1. Timeline of transitions between human operator and automation control. (Time increases from left to right.) The timeline indicates who is given authority to execute a skilled behavior to accomplish a particular task. Automation authority begins at an *initiation* event, and ends at a *termination* event.

promote operator comfort and system usability. In an environment where automation and human share responsibility, safety and performance are enhanced when both parties derive their actions and decisions from a common ground. Such a common ground can be established by creating automation controllers that a) choose control via perception-based control [2], [4] and b) present operator information in an ecologically-appropriate way [6], [9]–[12]. To this end, we create a computational framework of human-machine interaction using multiple dynamic mental model agents who coordinate skilled perception-based control mechanisms to generate satisficing[4] behavior. These dynamics include discrete switching as well as continuous control. They characterize the transitions between human and machine control as well as dynamics within the operator's cognition itself, but do not (in this paper) include learning and adaptation. They also characterize dynamic execution of skilled behavior as performed by the automation and the human operator.

Throughout the paper, we will make extensive use of equations and abbreviations. The following list summarizes the most frequently used notation:

1) KB, RB, and SB: knowledge base, rule base, and skill base;
2) SR, TR, BA: speed regulation, time-headway regulation, and active braking;
3) θ, u, c: state, action, and consequence;
4) μ_A, μ_L, S_b: accuracy metric, liability metric, and satisficing set.

II. Automation Taxonomy

Human interaction with automation follows the timeline diagrammed in Fig. 1. Initiation and termination of automation are functions of human desires and capabilities on the one hand, and machine design and capabilities on the other hand. Thus, the usefulness of an automation system is a function of both the operator's understanding of the dynamic machine-environment interaction as well as the operator's understanding of the automation. An automated system must facilitate not only seamless transitions between automated and human skills, but also unambiguous assignment of responsibility for switching between these skills. Effectual operator understanding is facilitated by designing automation systems compatible with human attentional, perceptual-motor, motivational, and cognitive characteristics.

[4]The term *satisfice* was used by Simon to denote the notion of being "good enough." The etymology of the word is unclear to the authors; some references state that it is an old Scottish word, and other references state that it is a composition of the words *satisfy* and *suffice*.

A. Skill Initiation

Sheridan identifies ten levels of automation in human-computer decision-making on a responsibility spectrum ranging from the operator deciding on a task and assigning it to the computer, to the computer deciding on a task and performing the task without input from the operator [16]. Based on these two extremes, automation that shares responsibility with a human operator can be broadly classified into two main categories:

1) Task Automation Systems: The operator chooses to delegate skilled task execution to the automation to relieve some physical or mental burden. Conventional and adaptive cruise control are examples of task automation systems.

2) Response Automation Systems: The automation preempts human decision making and control and initiates skilled task execution to facilitate safety or efficiency. Collision avoidance warnings and interventions are examples of response automation systems.

The essential distinction between these two categories is how the automation is *initiated* and, more precisely, who invokes the automation. In the first, the human operator initiates the automation whereas in the second, automation initiates itself.

B. Skill Execution

After the automation is initiated, it proceeds to perform its assigned skilled task execution. The manner in which the automation executes this skill is important since some execution schemes are more compatible with operator understanding than others. Moreover, if the task can be performed manually by the operator, then the operator may have preferences in how the task is executed; these preferences influence the operator's expectations for automated skilled task execution and therefore the transparency of the automation. It is desirable for the automation to perform the task in a way that conforms to the operator's expectations and preferences, and that is amenable to interruption. Remaining "in the loop" requires being attentive and having accurate and reliable expectations of the skilled-behavior of the automation.

C. Skill Termination

Automation will terminate if the assigned task is *completed* or if the human operator *intervenes*. Since *completion* and *intervention* can both occur, it is important to design task automation systems that help operators detect and respond to the limits of the automation. This observation leads to a second division among automation types. This division is exemplified by Sarter's automation policies [29].

1) Management by Exception: When operators can easily detect and respond to the limits of automation, then the automation, once initiated, is responsible for system behavior unless and until the operator detects an exception to nominal system behavior and terminates the automation. Examples of this termination policy include conventional and adaptive cruise control.

2) Management by Consent: When the limits of automated behaviors are not easily identifiable by operators, then the automation, once initiated, must convey its limits to the operator

and clearly indicate when it terminates itself. This allows the operator to develop accurate and reliable expectations of automation termination by consenting to a limited scope of automated behavior. Examples of this termination policy include timed devices and systems that perform a task with a clearly identifiable state of completion (e.g., lane change, parking, and emergency stopping).

The essential distinction between these two classes is how the automation is *terminated* and, more precisely, who turns off the automation. In the first (automation by exception), people terminate the automation whereas in the second (automation by consent) the automation terminates itself.

D. Transitions

While driving an automobile, many of us have had the following experiences:

1) We have been maintaining a safe but fast speed when suddenly traffic slows and we are forced to switch from a "speed regulation" behavior to a "stop and go" behavior. Such a switch is *mandated* by the change in the environment.
2) We have been contentedly following another vehicle for a long time when we suddenly decide that we should pass. We *discretionarily* choose to switch from a "car following" skill to an "overtaking" skill.

In the process of coordinating behavioral skills, there are *discretionary* switches between behaviors and *mandatory* switches between behaviors. Mandatory switches result from nature "forcing" a principal agent[5] (defined as either the operator or the automation) to change its behavior, and discretionary switches result when nature "affords" an alternate behavior which may be more desirable and the principal agent decides whether to change or not.

When a principal agent is operating in a natural environment, there will be times when both mandatory and discretionary behavior switches may occur. These behavior switches are examples of the more general *forced* and *suggested* behavior switches that emerge from the interaction between principal agents. When the operator is performing skilled task execution, the automation can *suggest* a change in behavior by generating a warning or informing the operator of a (in the automation's opinion) superior alternative, and the automation can *force* a change in behavior by intervening in system control or blocking operator actions. When the automation is in control, the operator can *suggest* a change in behavior by requesting a service or giving input, and the operator can *force* a change in behavior by terminating the automation or otherwise intervening.[6]

[5]Rather than using the term *agent*, we use the term *principal agent* to refer to the operator and the automation because the former term is used to refer to mental model agents.

[6]The distinction between mandatory and discretionary switches is also important in the design of decision support systems which form a significant component of high level task automation [20].

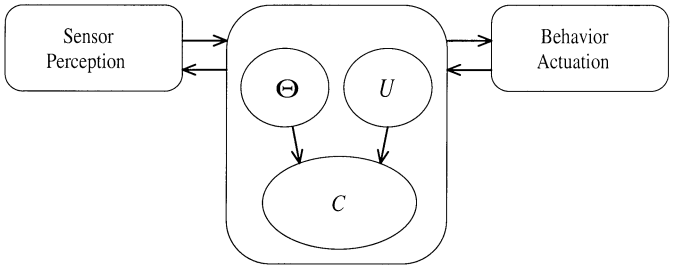


Fig. 2. Working specification of a mental model. In the figure, arrows mean influence so, for example, an arrow running from the sensor-perception block implies that the mental model affects perception while at the same time perception affects the mental model.

III. Model of Human Behavior Generation: Multiple Dynamic Mental Models

A human operator interprets and responds to sensory input according to the context established by a mental model[7] through task-specific filtering of the external world [17] and [21]. The operator must map environmental cues onto selected activities; an efficient way to perform this mapping is to employ a task-specific pattern of activities, and then implement these skills when appropriate (i.e., when "afforded" by the environment [6] and [24]). In this context, the term *activity* means the operator's actions on the system (e.g., an activity is pushing the brake pedal or turning the steering wheel), and the term *skill* means a learned sequence of operator activities. Within the overall context of goal-directed behavior generation, *tasks* can be defined and corresponding skills can be identified that may be used to accomplish the task (depending on conditions in the environment). Thus, behavior generation is organized into behavioral quanta that correspond to separate mental models[8] each with their own perceptually delineated operational domain. In this section, we present a framework of human behavior generation that can be used in the model-based design of human-centered automation.

A. Multiple Mental Models

A mental model is an internal mechanism employed to encode, predict, and evaluate the consequences of perceived and intended changes to the operator's current state within the dynamic environment. Formally, a mental model $\mathcal{M}$ is a triplet consisting of the following:

1) perceived state of the environment Θ;
2) set of decisions or actions U;
3) 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

[7]Note that some behaviors are describable by the stimulus-response model found in behaviorist literature. Such behaviors are described as *autonomous*, meaning that they are performed without any intervening cognitive processes, such as those associated with a mental model. We adopt the perspective in [25] which questions the existence of sophisticated autonomous behaviors and instead suggests that those behaviors which appear autonomous have some intervening cognitive influence, albeit a minimal influence because executing the behaviors is very efficient. Thus, the multiple mental model approach can be used as an abstraction of autonomous behaviors too.

[8]Our proposed model complements Wickens' model of information processing [31]. The model proposed herein can be viewed as an elaboration of how working memory, long-term memory, and response selection interact.

predicted consequence c, but also induces an evaluation of preferences (i.e., an ordering) among consequences (see Fig. 2). This mapping and evaluation of consequences depends on the driver's own mental and physical state, so the mental model subsumes these important elements of decision-making. The mental model $\mathcal{M}$ provides the context for meaningfully assessing sensory information and generating purposeful behavior.

Based on limited attentional resources in human operators, human behavior generation can be described using multiple mental models (treated as agents) which can be organized into a society of interacting agents [21]. This societal structure not only determines which agents contribute to operator behavior, but also which agents can employ attentional resources. A three-level [3], [28] multiresolutional society of interacting mental models organized into a hierarchical structure can be constructed corresponding to Rasmussen's knowledge-based (KB), rule-based (RB), and skill-based (SB) behaviors [26], [30]. 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 skilled task execution. Intuitively speaking, the KB, RB, and SB agents think, monitor, and control, respectively.

Each mental model $\mathcal{M}$ can be 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[9] whereby environmental information is actively perceived and interpreted, and when *disengaged* the mental model releases attention hence no such active perception occurs. In terms of Fig. 2, the mental model is enabled if the arcs between the mental model and behavior/actuation are active (hence, behavior u is actuated) and the mental model is engaged if the arcs between the mental model and sensor/perception are active (hence θ is actively perceived). Based on evidence that humans can monitor without controlling and control without attending, we suppose that $\mathcal{M}$ need not be enabled to be engaged, nor conversely. Switching from one skill to another (terminating one skill and initiating another) requires a fluency in enabling, disabling, engaging, and disengaging mental model agents. Depending on the adopted model for attentional limitations (e.g., single channel or limited resources), multiple mental models may (under conditions of limited resources) or may not (under conditions of single channel attention) be engaged.

Discussion of KB agents is an area of future research (and is more relevant to a discussion of cognitive decision aids than to a discussion of task and response automation). 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 managing SB agents and determining which 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. The role of an SB agent is to execute a *perception-based control law* that performs the task specified by the RB agent. The remainder of this section elaborates on how RB and SB agents can be modeled using cognitively plausible descriptions of decision-making and perceptual-feasible descriptions of control, respectively.

B. Rules: Skill Transitions

Since each task may have multiple skills that can accomplish the task for different environmental circumstances, successful task completion requires skills to be managed. Satisficing decision theory (SDT) [13], [14], which employs and compares two evaluation functions similar to the way benefit and cost are compared in economics literature, is an ideal tool to describe operator management of multiple skills. In SDT, preferences over consequences are partitioned 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*. In our mental model context, recall that U denotes the set of possible decisions or actions and Θ denotes the set of possible perceptual states.[10] For each decision $u \in U$ and for each perceptual state $\theta \in \Theta$, a consequence results which is the effect of making decision u when nature is in state θ. The accuracy μ_A: $U \times \Theta \mapsto \Re$ and liability μ_L: $U \times \Theta \mapsto \Re$ functions are preference relations defined for each consequence (i.e., action/state-of-nature pair). All consequences which are more accurate than liable are acceptable [14], [15], from whence we can define the satisficing set as $S_b = \{(u,\theta)\colon \mu_A(u;\theta) \geq b\mu_L(u;\theta)\}$ where b is a design parameter that we leave unspecified here but identify in experiments.

Given SDT, we can restrict attention to those states which are satisficing for a given u, and those skills which are satisficing given the state of nature, respectively defined as $S_b(u) = \{\theta\colon \mu_A(u,\theta) \geq b\mu_L(u,\theta)\}$ and $S_b(\theta) = \{u\colon \mu_A(u,\theta) \geq b\mu_L(u,\theta)\}$. In terms of behavior management by an operator, suppose a skill $u \in U$ is being used to produce behavior. The operator monitors θ, and when $\theta \in S_b(u)$ no change is necessary. However, when $\theta \notin S_b(u)$, the current behavior is not acceptable and must be terminated, and a behavior that is appropriate for the circumstances must be initiated. Given the mandatory need to terminate or the discretionary choice to terminate a skilled behavior u, any skilled behavior $u' \in S_b(\theta)$, including automated behaviors, can be initiated.

C. Skills: Perception-Based Dynamics and Control

Skills can be emulated by closed-loop controllers that operate on environmental cues. These cues must be perceptually plausible meaning that operators must be able to sense them. For example, automobiles drivers performing the speed management task can use vehicle speed v_A, time headway T_h, and time to collision (or, equivalently, inverse time to collision T_c^{-1}) as perceptually feasible perceptual cues [10].

[9]Attentional sharing is necessary because operators have limited computational and memory resources. A simple attentional model schedules attention between agents. More realistic models for attention are an area of future research. We refer the reader to [3], [8] for more about the role of attention in this process.

[10]A *perceptual state* is more general than a *perceived state*. A perceived state includes those ecologically valid states that are immediately observable from the environment, but a perceptual state includes both immediately observable ecological states as well as some history of what has been observed in the past.

System dynamics can be represented using classical Newtonian dynamics with a corresponding dynamic state vector $\boldsymbol{x}$ that, depending upon the actions of the operator A and the actions of a disturbance B (denoted u_A and u_B, respectively), can be described as a discrete time dynamical system to describe how the state $\boldsymbol{x}$ changes over time (indexed by k) yielding $\boldsymbol{x}_{k+1} = \boldsymbol{g}(\boldsymbol{x}_k, u_A, u_B)$. Shifting focus from a world centered (i.e., Newtonian physics) perspective to a human-centered perspective (i.e., perception-based dynamics), we construct a model of skilled behavior using a discrete time dynamical state space representation that possesses the following five desirable features:

- Feature 1: perceptual state variables θ, possibly different from $\boldsymbol{x}$, are perceivable by the operator;
- Feature 2: the space spanned by θ (denoted $\mathrm{sp}(\theta)$) equals $\mathrm{sp}(\boldsymbol{x})$;
- Feature 3: an internal dynamical model of perceptual state transitions $\theta_{k+1} = \boldsymbol{f}(\theta_k, u_A, u_B)$ can be constructed ($\boldsymbol{f}$ denotes the dynamical response in space θ, and $\boldsymbol{g}$ denotes the related dynamical response in space $\boldsymbol{x}$);
- Feature 4: a control law $u_A = \pi(\boldsymbol{f}, \theta)$ can be constructed from the internal model and the observed perceptual state using cognitively plausible decision mechanisms;
- Feature 5: decision planes can be described in a low dimensional subspace of $\mathrm{sp}(\theta)$ (i.e., decisions depend on relatively few variables).

These five desirable features are satisfied by the multiple mental model framework. This framework satisfies Features 1–3 because correct mental models are perceptual state-based representations of how consequences evolve in the real world. Furthermore, the multiple mental models perspective allows some mental models to act at the skill level (thereby satisficing Feature 4) and others to act at the rule level (thereby satisficing Feature 5).

D. Relevance to Task Automation

By restricting attention to task automation, we limit the set of problems that we address to those in which the operator discretionarily initiates the automation; see [7], [9] for a discussion of human-centered response automation. In the remainder of this paper, we focus attention on task execution and task termination. In this section, we relate the multiple mental model framework to the task automation problem. We begin by discussing how skill management impacts the termination policy of task automation, and then discuss how the execution scheme impacts the comfort and usability of the automation.

1) Skill Management: Let u_{AUTO} represent the skilled behavior executed by the automation. Since the automation has limitations, there exist a set of environmental states θ that support automation (i.e., that are within the operational limits of the automation); thus, we can identify the set $S_b(u_{\mathrm{AUTO}}) = \{\theta\colon u_{\mathrm{AUTO}} \text{ is acceptable}\}$. Given an environmental state θ and skilled operator behavior u, the operator can discretionarily terminate u and initiate the automation only if $u_{\mathrm{AUTO}} \in S_b(\theta)$.

In making the transition from automated to manual behavior, a switch is mandated when $\theta \notin S_b(u_{\mathrm{AUTO}})$. Since operators are situated in the environment and thereby experience a dynamic system, the state θ is dynamic. The transition from $\theta \in S_b(u_{\mathrm{AUTO}})$ to $\theta' \notin S_b(u_{\mathrm{AUTO}})$ is the *perceptual triggering event* that mandates automation termination. This perceptual triggering event is naturally detected in manual operation provided[11] that the appropriate perceptual cues are receiving attention.

We are now in a position to state the first hypothesis for desirable human-centered automation.

2) Hypothesis 1: Mandatory Termination of Task Automation: Assuming θ receives attention, switches from automated to operator control via operator intervention are easiest for the operator if there exists a $G \subset U$ such that $S_b(u_{\mathrm{AUTO}}) = \cup_{u \in G} S_b(u)$. In words, if the limits of automation correspond to the limits of a subset of natural operator skills, then the limits of the automation are most likely to be perceived and detected by the operator.

The motivation for this hypothesis is related to the multiple mental model framework outlined in previous sections. The operator has an understanding of how his or her own skills should be used (i.e., experienced operators are good at managing their own skills) and, more precisely, when and how mandatory termination should occur.[12] Task automation then becomes a means wherein the operator delegates to the automation the responsibility for actuating the skill. The operator thus disables their own perceptual skills (i.e., an SB agent is disabled), but continues to monitor the world hence the operator's appropriate mental models are engaged (i.e., a disabled SB agent is engaged). Thus, the operator is able to attend to relevant perceptual cues (although they may not do so) and, using these cues, assess and understand the limits of the automation. This understanding translates into accurate and reliable expectations of the automation's abilities, and these expectations facilitate effective automation termination under the management by exception policy. In essence, the hypothesis states that when the automation matches a subset of the operator's skill set then the operator can detect the operational limits of the automation.

3) Skill Execution: The second hypothesis is relevant when either the operator needs to estimate the current system state, such as when the system may need to be interrupted, or when the operator experiences physical dynamics as a result of executing a skill (e.g., feeling forces when an automobile decelerates). Since operators use perceptual skills (or other afforded behaviors) to operate the system, then the system should emulate the operator in some sense if the system is to supplant operator

[11]There 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 assessable description of the limits of system behavior. For experienced operators, most cognitive mistakes are characterized by lack of attention rather than the ability to assess perceptual triggering events.

[12]Mandatory termination does not always occur when it should. Under conditions of stress, fatigue, etc., humans may experience cognitive tunnel vision wherein a single task or limited set of tasks consumes all cognitive resources.

control. Design of such emulation-based automation can help ensure that the operator is both comfortable with system operation and confident in his or her ability to efficiently and safely terminate the automation.

4) Hypothesis 2: Dynamics of Task Execution: Assuming that a particular skill is appropriate for a given context, operators are more likely to understand and prefer automation controllers that emulate human perception-based dynamics.

Stated simply, if operators perform a skill correctly and efficiently then the system dynamics should mimic the operator. Doing so helps the operator develop reasonable expectations and/or allows the operator to experience comfortable system dynamics.

The motivation for this hypothesis is related to the nature of human perception. We assume that either the operator has experience in performing the skill or will experience physical forces when the skill is performed. In either case, the operator generates expectations about how the automation will behave and how such behavior will impact the operator. If these expectations are difficult to generate, then the operator will perceive the automation as either a capricious companion or a companion that should not be questioned. No matter which opinion of the automation is held, it is clear that mismatched expectations will defeat the purpose of human-machine cooperation [22]. Given that expectations are so important, we note that it can be difficult, or impossible, to generate accurate expectations if the operator cannot associate environmental cues that are perceivable by humans with the actual automation behavior. Human-centered design is based in this observation; it is important for automation to respect how humans perceive the environment and act in a way that facilitates reasonable expectations. These expectations are best managed when the operating range of the automation is aligned with the human operator, and when automated execution of the behavior matches skilled human behavior.

IV. Automobile Drivers and ACC Design

Our approach to describing, predicting, and enhancing driver behavior is to identify the set of skills that automobile drivers use to safely manage speed and interact with traffic. To manage speed and interact with traffic, we suppose that drivers use a set of learned skills [19]. Our approach then identifies how one skill-based behavior is switched to another and how perceptual cues trigger such switches. This approach produces a computational model that emulates driver behavior and, by associating ACC behaviors with a subset of natural driver skills, can be extended to predict how the driver switches between manual and automated behaviors by detecting and interpreting the operational limits of the automation. These predictions are then supported by experimental results and evidence gathered from relevant literature.

A. Automobile Driving Example

Longitudinal control includes three closed-loop perception-based controllers, only one of which is enabled at a time: 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.[13] Thus, the set of skills relevant to longitudinal vehicle control (i.e., speed management) are $U = \{\mathrm{TR}, \mathrm{SR}, \mathrm{BA}\}$. The activities associated with these skills are pressing the brake or accelerator pedal.

B. Transitions Between ACC and Driver Control

Since we have restricted attention to task automation systems, we are primarily interested in how the driver can sense when ACC has reached its operational limits and thus terminates ACC control. It is useful to associate ACC functions with a subset of the human driving skills described in the previous section. In the absence of other traffic, an ACC system regulates speed about a preset value and thereby, automates $u = \mathrm{SR}$, meaning the enabled skill u is **s**peed **r**egulation. In the presence of other traffic, an ACC system regulates time headway about a preset value and thereby automates $u = \mathrm{TR}$ (**t**ime headway **r**egulation). The transition between these skills, including active braking, is a critical aspect of ACC usability. Two alternative methods for such transitions are of importance: engine braking and active braking. We argue that the active braking skill is distinct from engine braking, where the former is limited to the BA (**b**raking **a**ctive) skill and the latter is used in the TR skill.

1) Hypothesis 1: Based on the assumed perspective that ACC systems are designed to safely increase comfort, we are primarily interested in when θ is such that the ACC is not satisficing, $u_{\mathrm{ACC}} \notin S_b(\theta)$. Such an event can occur if either the ACC system malfunctions or the state of the environment is outside of the scope of the ACC system. Focusing on the second occurrence (we will assume that the first occurrence is negligible—an assumption that must be considered in practice) our task is to determine the perceptual trigger between satisficing ACC behavior and unacceptable ACC behavior. Since the limits of ACC behavior as a function of traffic, weather, time of day, and infrastructure correspond to bounds of $S_b(u_{\mathrm{ACC}})$, this task translates into detecting and interpreting the bounds of the satisficing set. Hypothesis 1 becomes:

2) Hypothesis 1: Driver Intervention in Task Automation: Assuming θ receives attention, switches from ACC to driver control via driver intervention (i.e., forcing ACC to terminate) are easiest for the driver if $S_b(u_{\mathrm{ACC}}) = S_b(\mathrm{TR}) \cup S_b(\mathrm{SR})$.

In other words, an ACC system that automates both speed and time headway regulation is most likely to facilitate attentionally manageable and seamless transitions from automation to human control. (Note that BA must be performed by the human which means that the brake is pushed. We restrict attention to systems where pushing the brakes turns the ACC off, which means that the driver must turn it back on before the ACC will again affect vehicle behavior.) If Hypothesis 1 holds, then the set of existing perceptual cues θ used by the driver to detect mandatory switches to active braking can also be used to detect when the ACC system should be disengaged (hence a management by ex-

[13]We do not consider alternative collision avoidance strategies such as swerving because these strategies emerge from the interaction between multiple RB agents. Instead, we leave these areas for future work on the fusion of RB agent behaviors.

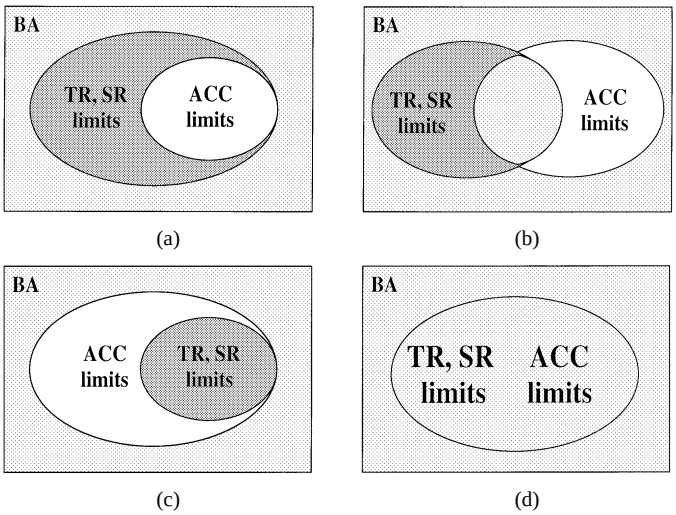


Fig. 3. Comparison of TR/SR domains and ACC domain. (a) TR/SR domain broader than ACC domain $S_b(u_{\mathrm{ACC}}) \subset S_b(\mathrm{TR}) \cup S_b(\mathrm{SR})$. (b) TR/SR domain and ACC domain incompatible $S_b(u_{\mathrm{ACC}}) \not\subseteq S_b(\mathrm{TR}) \cup S_b(\mathrm{SR})$. (c) ACC domain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

ception automation policy is acceptable). If Hypothesis 1 is violated, then drivers require either training or a surrogate system for detecting a mandatory switch (hence, a management by consent automation policy should be used). Note that this requires knowledge of a driver's subjective perceptual boundaries beyond which they actively press the brake (i.e., detection of the perceptual triggering event). As reported in subsequent sections, these boundaries are determined by experiment.

3) Alternatives to Hypothesis 1: In Figs. 3(a)–(d) the support for satisficing ACC behavior is shown in relation to the support for TR, SR, and BA behaviors for three idealized cases. Compare each of the cases in Figs. 3(a)–(c) to Hypothesis 1 in Fig. 3(d) wherein the sets overlap. For Fig. 3(a), the ACC system does not accomplish its stated objective of automating car following and speed regulation behaviors and, consequently, is neither successful from a designer's perspective nor useful from a driver's perspective.

For Fig. 3(b) the set of states for which ACC and TR/SR are applicable are incompatible. Such a design can either make it difficult for drivers to intervene in ACC control, or require either a period of driver adaptation to learn the limits of the new system or the inclusion of a surrogate system to indicate the limits of the ACC. For a useful system with a wide range of drivers, it is undesirable to design the ACC system that requires the driver to extensively learn and carefully monitor the automation to produce safe automation.

For Fig. 3(c), the ACC system exceeds the driver's capabilities. Such a system appears attractive in that more driver behaviors than just TR and SR are automated. The problem with this approach is that the ACC system does not automate all of the BA skill. By contrast to Fig. 3(d) in which a driver knows when to brake because the driver has clearly defined boundaries between TR/SR and BA skills, a driver has no such experience in detecting the limits of ACC behavior. Unless the limits of this behavior are easily perceived by the driver, such a system can result in an unsafe ACC design.

4) Implications From Hypothesis 1: ACC automates two skills normally performed by automobile drivers: speed regulation and nominal car following. This leaves braking to avoid collisions and other emergency behaviors to the driver. When the behavior of an ACC system exceeds the support for SR/TR skills how do drivers know when to intervene? Provided that drivers can attend to the perceptual cues, two methods for detecting the need to intervene are possible. First, drivers can be trained to learn the perceptual boundary and second, a surrogate can be used to help the drivers detect the boundary. Since the people who drive vehicles come from diverse backgrounds with diverse skills and training, it is unlikely that training will be universally effective. The second option is to use a surrogate to assist drivers in detecting the boundary. This not only helps drivers detect the need to intervene in ACC control, it also acts to train drivers regarding the limits of ACC behavior. However, the design of such a surrogate is a nontrivial task as demonstrated by the difficulty of designing a useful warning system. Such warning systems must be designed with careful attention to driver perceptual and information-processing capabilities. However, a carefully designed surrogate might increase the safety of any ACC system since Hypothesis 1 is an ideal and difficult to reach in practice. Unfortunately, a surrogate warning signal that is not aligned with a human's natural mental model boundaries may cause a human to adopt "wait and see" strategy wherein the human waits to intervene until it is clear that the automation will fail; this can cause delay in the transition between automation and manual control, and this is undesirable in safety critical tasks.

C. Automation Dynamics

Since ACC systems produce physical forces on the operator and require the operator to stay "in the loop", the way an ACC system behaves is important. A useful method for assuring that ACC performance matches human expectations and produces comfortable execution dynamics is to have the ACC emulate human operator behavior.

1) Hypothesis 2: Again based on the assumed perspective that ACC systems are designed to safely increase comfort, we recognize that ACC systems should be compatible with human perceptual capabilities and desired comfort. Since automobile drivers are experienced in regulating speed and time headway, they have reasonable expectations about the performance of automation. These expectations translate into the statement of Hypothesis 2 as the following:

2) Hypothesis 2: Dynamics of ACC Behavior: Drivers understand and prefer ACC controllers that emulate human perception-based dynamic control.

Because driver's experience dynamical forces when an ACC system operates, the ACC needs to be comfortable. This will require some good signal processing and may even require a new class of sensors. The alternative to this hypothesis is for engineers to design ACC systems that operate in traditional ways and run the risk of being uncomfortable or unsafe for the average driver. We discuss the consequences of such designs in the next section.

3) Alternatives to Hypothesis 2: The alternatives to Hypothesis 2 include requiring humans to develop new expectations

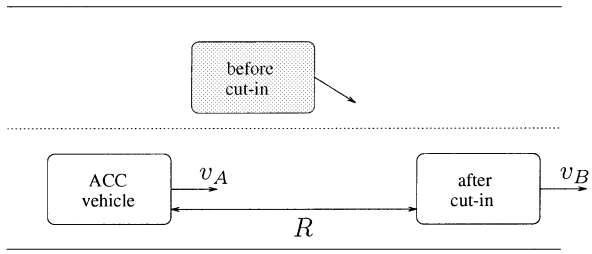


Fig. 4. "Cutting in" problem. The lead vehicle prior and subsequent to the cut-in event is represented by a shaded box above the ACC vehicle, and an open box in front of the ACC vehicle, respectively.

of ACC behaviors, or adding information systems that help human's understand ACC behaviors. The latter approach requires the same careful analysis of human perceptual abilities, and will likely shift the operator burden from physical control to conscious monitoring. Such a shift can produce ironies of automation and can decrease system safety [1]. The former approach assumes that humans can almost universally learn how a new system behaves and adjust their thinking to accommodate this behavior. Not only is this risky from a liability perspective, it also assumes a great deal about the intended users of the ACC. It is unlikely that human drivers will be able to suppress past experiences and always successfully interact with poorly designed ACC systems, especially in safety critical situations.

4) Implications of Hypothesis 2: The implications of Hypothesis 2 are simple: either translate the vehicle's raw sensor data into a perceptual domain and then design a controller for this domain, or retain a state space description in the sensor domain and then design a controller that emulates human braking and acceleration.

V. Experimental Evidence

In this section, we present not only evidence that supports the multiple mental model framework, but also evidence that supports the two hypotheses. Included in this evidence are results from relevant literature that supports the human-machine interaction hypotheses.

A. Multiple Mental Models: Experimental Support for Framework

To test Hypothesis 1, we have gathered evidence from the transition between speed regulation and active braking. We present this evidence in this section.

1) Experiment I: Behavioral Quanta and Skill Switches: To determine SDT-based models of driver behavior, we will focus on the "cutting in" problem wherein vehicle B cuts in front of the driver's vehicle (vehicle A) as diagrammed in Fig. 4. Subsequent to a cut-in event, we refer to the vehicle that cuts in as the lead or cut-in vehicle. In the figure, v_A and v_B represent the velocities of the driver's vehicle and the lead vehicle, respectively, $v_R = (dR/dt) = v_B - v_A$ represents the relative velocity between the vehicles, and R represents the range (relative distance) between the vehicles. We define time headway and inverse time to collision, respectively, as $T_h = (R/v_A)$ and $T_c^{-1} = -((v_R)/R)$, where v_A is the speed of the driver's vehicle, R is the relative distance (range) between the driver's vehicle and a lead vehicle, and v_R is the rate of range change (relative velocity between the driver's vehicle and the lead vehicle). Given these perceptual values, a perceptual state can be defined as $\theta = [T_c^{-1}, T_h, v_A]^T$.

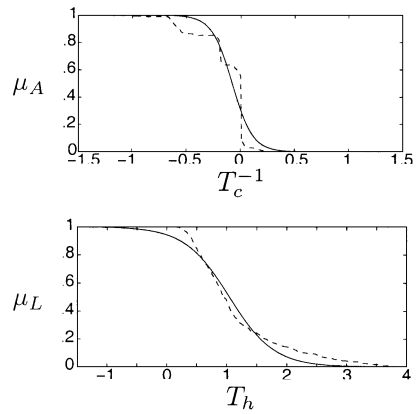


Fig. 5. Actual (dashed line) and approximated (solid line) membership functions: (a) accuracy $(\alpha = -32.5, \beta = -0.04)$ and (b) liability $(\alpha = -4.9, \beta = 1.10)$.

a) Description: Automobile driving is a mix of cognitive and behavioral skills. When a driver delegates a task to automation, the vehicle assumes responsibility for a behavioral skill. However, the driver retains (meta) responsibility for detecting the limits of the automation and responding appropriately. We conducted an experiment in which human subjects were placed in a driving simulator[14] with a cruise control system engaged. At random intervals, a vehicle cut in front of the subject's vehicle and compelled the subject to determine if the automation can safely perform the skill or if the driver needed to intervene. Empirical estimates of accuracy and liability can be obtained as described below. Fig. 5 presents the resulting empirical estimates and the best fit curve to these estimates. Note that, for this example, the ecologically valid state variables $\theta = [T_h, T_c^{-1}, v]$, (time headway, inverse time-to-collision, and velocity, respectively) suffice to describe the domain of expertise Θ [8]. Data were partitioned into two classes: *active braking* (brake pedal depressed) and *nominal behavior* (CC engaged, accelerator depressed, or engine braking).

b) Empirical Estimates: To identify μ_A and μ_L, our objective is to find substates that trigger active braking. We therefore distinguish between nominal behavior $u \in \{\mathrm{SR}, \mathrm{TR}\}$ and active braking behavior $u = \mathrm{BA}$. Our goal is thus to find when $\theta \notin S_b(u)$ for $u \in \{\mathrm{SR}, \mathrm{TR}\}$. Nominal operating conditions occur when the brake pedal is not pressed. For both nominal and braking conditions, we select representative sample points from each experimental trial and create two sets of $[T_c^{-1}, T_h]^T$ points: one set for nominal conditions, denoted NOM, and one set for braking conditions, denoted BRK. For trials when subjects actively brake, the sub-state(s) $[T_c^{-1}, T_h]^T$ when braking is initiated is included in BRK, and the sub-state(s) $[T_c^{-1}, T_h]^T$

[14]Nissan's SIRCA simulated driving environment includes approximately six miles of highway with three lanes in each direction and ambient traffic. In an experiment using the SIRCA environment, a subject performs lateral control but engages a cruise control (CC) system to perform longitudinal control about a preset condition ($v^* = 20$ m/s ≈ 43 mph). During the experiment, a cut-in vehicle passes the subject's vehicle while the CC is engaged and cuts into the lane with a specified relative velocity $v_R(0)$ and randomly selected initial time headway $T_h(0)$.

when braking is terminated is included in NOM; for trials when subjects do not brake, the initial sub-state $[T_c^{-1}, T_h]^T$ in the trial is included in NOM; and for trials where subjects only brake (by anticipating the cut-in and then coming to a stop), the initial sub-state $[T_c^{-1}, T_h]^T$ in the trial is included in BRK.

For notational purposes in the subsequent sections, let $N(T = \tau \,|\, \text{CONDITION})$ denote the cardinality of the set of points $T = \tau$ given CONDITION. For example, $N(T_c^{-1} = \tau | \text{NOM})$ is the number of points in the set $\{\theta \in \text{NOM}: T_c^{-1} = \tau\}$. Under nominal conditions ($\theta \in$ NOM), relative velocity must be considered acceptable to the driver hence the distribution of T_c^{-1} under nominal conditions is an observable entity that provides information about what is accurate. Clearly, if $T_c^{-1} = \tau_2$ is accurate, then $\tau_1 < \tau_2$ must be at least as accurate. This monotonicity property facilitates the computation of the accuracy function as the cumulative distribution function

$$\begin{aligned}\mu_A\left(T_c^{-1} = \tau\right) &= 1 - F_{T_c^{-1}}(\tau \,|\, \text{NOM}) \\ &= 1 - \frac{N\left(T_c^{-1} \le \tau \,\middle|\, \text{NOM}\right)}{N\left(T_c^{-1} \le \infty \,\middle|\, \text{NOM}\right)}.\end{aligned}$$

For classification purposes, we fit (via least squares) a sigma function of the form $1/e^{(-\alpha(\tau-\beta))}$ to $\mu_A(\cdot)$ yielding the function shown in Fig. 5(a).

When braking is initiated ($\theta \in$ BRK), time headway values must be considered unacceptable hence the distribution of time headways when the driver initiates braking is an observable entity that provides information about what is rejectable. Clearly, if $T_h = \tau_2$ is rejectable then $\tau_1 < \tau_2$ must be at least as rejectable. This monotonicity property facilitates the computation of the liability function as the cumulative distribution function

$$\begin{aligned}\mu_L(T_h = \tau) &= 1 - F_{T_h}(\tau | \text{BRK}) \\ &= 1 - \frac{N(T_h \le \tau \,|\, \text{BRK})}{N(T_h \le \infty \,|\, \text{BRK})}.\end{aligned}$$

For classification purposes, we fit (via least squares) a sigma function of the form $1/e^{(-\alpha(\tau-\beta))}$ to $\mu_L(\cdot)$ yielding the function shown in Fig. 5(b).

c) Classification Results: For the driver to switch from one skill to another, it is necessary to identify when $u \notin S_b(\theta)$. Using $\mu_A(T_c^{-1})$ and $\mu_L(T_h)$ from Fig. 5, we can construct the set of states $\mathcal{S}_b = \{\theta\colon \mu_A(T_c^{-1}) \ge b\mu_L(T_h)\}$ that support nominal behavior, and the set of states $\mathcal{S}_b^c = \{\theta\colon \mu_A(T_c^{-1}) < b\mu_L(T_h)\}$ (superscript c denotes complement) that do not support nominal behavior. If $u \in \{\text{TR}, \text{SR}\}$ and $\theta \in \mathcal{S}_b^c$ then $\theta \notin S_b(u)$. Thus, the line $\mu_A(T_c^{-1}) = b\mu_L(T_h)$ determines when behavior must be switched from nominal to braking. In other words, the line represents the boundary of $S_b(\text{SR})$.

Given the empirically derived membership functions, we can determine the boundary between nominal and braking behaviors as a function of b by finding the perceptual states θ for which $\mu_A(T_c^{-1}) = b\mu_L(T_h)$. This is illustrated in Fig. 6 for the data gathered in the simulator experiment, where $\circ$ indicates $\theta \in$ NOM and $\times$ indicates $\theta \in$ BRK. To the northwest of the line, BA is satisficing but TR and SR are not, and to the southeast of the line TR and SR (and, perhaps, BA) are satisficing.

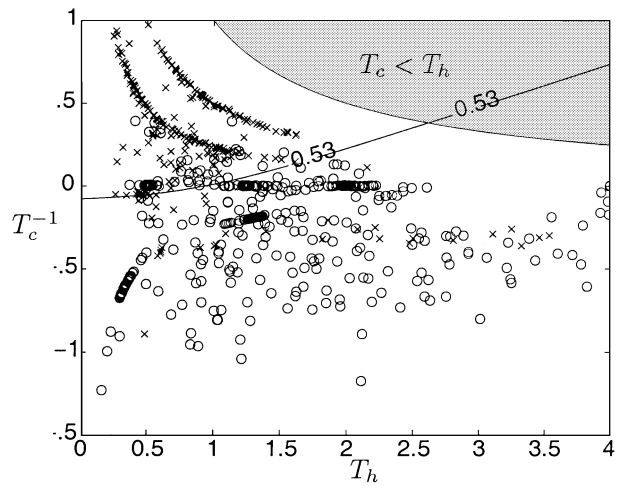


Fig. 6. Scatter plot of nominal and braking perceptual states.

Classification can be performed by finding the value of b that optimally separates braking from nominal behavior. The value $b = 0.53$ is the minimax value over percent misclassifications, percent false alarms, and percent missed detections and indicates that over 85% of samples are correctly classified.

2) Experiment II: Behavioral Quanta and Skill Switches: Because Experiment I relied on a fixed-base driving simulator, there is some question about how these results relate to situated driving in real vehicles. To test the transfer of these results to driving, a second experiment was conducted with professional drivers responding to unpredictable cut-in events with real vehicles on a test track. In the experiment, two vehicles drive in the same lane on a closed test track. The subject drives vehicle A which follows vehicle B. The drivers in vehicles A and B are required to maintain an assigned speed $v_A(0)$ and v_B until a chime rings in vehicle A's car. When the chime rings, the driver of vehicle A is to establish a natural following distance (i.e., drive as if vehicle B had just cut-in to vehicle A's lane) while vehicle B maintains a constant speed. Measurements include R, v_A, brake pressure, and throttle opening angle. Time headway and time to collision were computed from these measurements. A complete description of this experiment can be found in [10]. To perform the classification, $\mu_A(T_c^{-1})$ and $\mu_L(T_h)$ were estimated, and the b that minimizes the misclassification error was determined. The results indicate one false alarm and no missed detections.

B. Transitions: Experimental Support for Hypothesis 1

We now turn attention to experimental evidence that indicates what is an acceptable termination policy, and what is not acceptable.

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

(see Section V.A and [8], [12]). 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 drivers must continue to steer the vehicle. This 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.

2) Advanced Cruise Control: Consider the results from [23] wherein an ACC system that included a limited amount of active braking was studied. Among other things, the study reported that ACC users have "too large (of) expectations" about ACC abilities, most ACC users waited to intervene until a warning was received, and collisions that occurred (in the driving simulator) when the ACC was engaged "could (not) be explained . . . by decreased level of driver alertness" (which may not hold when drivers are not engaged in an experimental study). The first two of these findings suggest that when an ACC system does more than automate TR and SR skills, then drivers have a difficult time safely detecting the limits of the ACC system (there is no natural switch between skills). Instead, drivers sometimes adopt a "wait and see attitude" that allows the ACC system to reach its limits rather than proactively intervening to avoid an unsafe situation. The last finding suggests that the negative effects of ACC behavior cannot be entirely attributed to not attending to θ.

C. Dynamics: Experimental Support for Hypothesis 2

We now present evidence that suggests that when Hypothesis 2 is violated, the resulting performance of an ACC system is unacceptable or unnatural. Additionally, this evidence supports the claim that people use perceptual cues to perform manual control in automobile driving. We refer the reader to other work on perception-based control in [4], [6], [10], [18] for further evidence, including a correlation analysis and nonlinear regression analysis of the variables used to control braking.

1) Description: Two elementary PD controller were developed and tested in various cut-in scenarios. These PD controllers operated on the error between estimated range R and desired range $R^* = v_A T_h^*$, where computation of R^* from v_A is a concession to technological constraints on estimating v_B. Instrumented vehicle were then equipped with the controllers, and the behaviors of the controllers were compared to the behaviors of professional drivers who were placed in similar situations using the experimental set-up described in Section V.A2.

2) Results: The data are classified into two categories: those for which active braking occurs and those for which no such braking occurs, indicated in Figs. 7–9 with a $\times$ and a $\circ$, respectively. The sequence of $\circ$'s present after the cut-in event indicate the amount of time taken to react to the cut-in event. It is helpful to illustrate the perceptual phase plane trajectories subsequent to the cut-in event for each data class (time histories can be found in [10]).

From the experimental data, four observations are apparent and worth noting. First, the decision to brake is made (ignoring reaction time) when $T_c^{-1} > 0$ ($v_R < 0$). Conversely, a driver is likely to accelerate when $T_c^{-1} < 0$ ($v_R > 0$). Thus, dividing driver behavior into active braking and nominal (not-active) braking produces a division roughly at $T_C^{-1} = 0$ ($v_R = 0$). Second, when $T_c^{-1} \leq 0$, the factor determining dynamic driver behavior appears to be related to time headway. This is observable from the driver response in Fig. 7 wherein the driver first establishes zero relative velocity (and infinite time to collision) and then appears to regulate vehicle speed around the time headway value $T_h^* \approx 1.65$ s. Third, drivers establish steady-state (i.e., $v_R \approx 0$) behavior at different values of T_h^*. There is a significant ($P \approx 2 \times 10^{-7}$) T_h^* difference between subjects who participated in the experiment. For subject A the mean terminal headway is $T_h^* = 1.47$, and for subject B the mean terminal headway is $T_h^* = 2.01$. Interestingly, there are no significant within subject T_h^* differences for different $v_A(0)$ or v_B conditions. Thus, we find evidence that T_h influences braking dynamics independently of v_A. Our conclusion is that

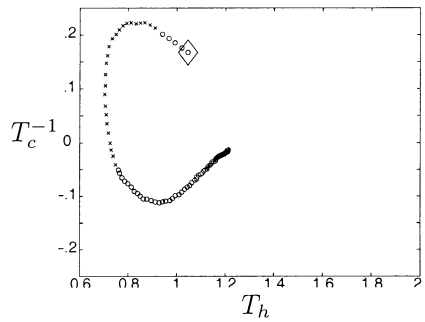


Fig. 7. Perceptual phase plane trajectories for professional driver one with initial conditions $(T_h(0), v_R(0)) \approx (1\ \mathrm{s}, 20\ \mathrm{km/hr})$.

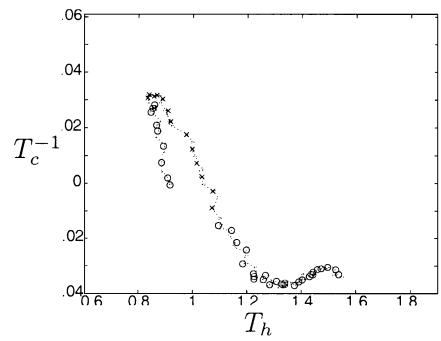


Fig. 8. Perceptual phase trajectory of unacceptable automated performance with initial conditions $(T_h(0), v_R(0)) \approx (1\ \mathrm{s}, 0\ \mathrm{km/hr})$.

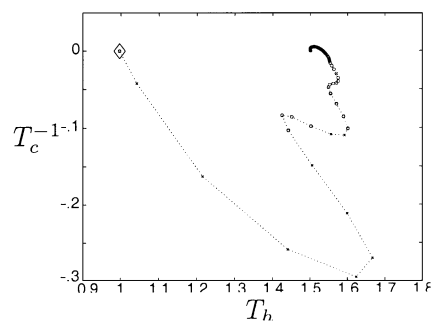


Fig. 9. Perceptual phase trajectory of unacceptable automated performance with initial conditions $(T_h(0), v_R(0)) \approx (1\ \mathrm{s}, 60\ \mathrm{km/hr})$.

driver's use time headway and time to collision as speed-independent variables for controlling the dynamics of braking as well as the set point for car following. Fourth, the characteristics of the phase plane trajectory influence the acceptability of the automated performance. As shown in Fig. 7, people establish natural following distances by generating *counterclockwise* trajectories in perceptual space. By contrast, automated controllers that fail to follow this counterclockwise trajectory are judged unacceptable by subjects. In the data shown in Fig. 8, the test driver reported that the performance was unacceptable because the braking action was too extreme given that the relative velocity, though positive, was small in magnitude. This "hard-braking plus low relative velocity" characteristic is manifest as an unacceptable clockwise trajectory in the perceptual phase plane. Additionally, as shown in Fig. 9, not only must perceptual trajectories be counterclockwise they must also be *smooth*. When trajectories are not smooth, subjects report a period of rapid braking followed by a coasting interval which was then followed by another braking interval. This surge, though moderate, reportedly felt awkward and unnatural and is unacceptable.

VI. Conclusions and Future Work

A human centered approach to designing automation is introduced and motivated through a case study of automated car following systems. The fundamental design principle is to use the human operator as a template for automation. A model of how operators perform various tasks is required if a template-based approach to human-centered design is to be successful. This model can be used to identify sub-components of a particular task that are characterized by a region in perceptual state space and a corresponding behavioral skill. As the perceptual state (e.g., vehicle, environment, and/or traffic) changes from one skill domain to another, operators switch skills. These skills are characterized by

1) perceptual variables used to perform skilled perceptual motor control;
2) perceptual variables used to decide when to initiate and terminate the skill;
3) understanding of how operators evaluate performance;
4) understanding of what control strategy operators adopt.

For example, drivers switch from speed regulation skill to car following skill depending on the time headway and time to collision to a preceding vehicle. It is hypothesized that automation of these easily identifiable sub-components has several advantages over approaches based on other criteria (e.g., technological feasibility). We provide experimental support for the hypotheses that such a human centered approach

1) improves detection of nonsupported situations;
2) improves evaluation of system performance;
3) facilitates development of a proper level of trust;
4) improves ability to take over control;
5) theoretically enhances safety.

Skilled operator behavior can be organized into behavioral quanta that correspond to separate skill-based behaviors. For longitudinal control of an automobile, such skills include following a lead vehicle, braking to avoid a collision, and regulating speed about a desired value. When automation is added to a vehicle, some of these skill-based behaviors are performed automatically by the vehicle itself. By measuring human subjects' responses to cut-in events in both driving simulator and test track studies, we have presented experimental support that natural boundaries exist between automated speed regulation (conventional cruise control) and skill-based braking behaviors. These experimental results are described in the theoretical and computational framework provided by using satisficing decision theory to describe switches between multiple mental models. Extending these results to ACC systems we hypothesize that, assuming an attentive driver, switches from ACC to driver control (via driver intervention) are easiest if the operational limits of ACC behavior correspond to the natural boundaries between that of the *speed-regulation/car-following* skill-based behaviors and that of the *active braking* skill-based behavior. We conclude that advanced vehicle system design can benefit from careful analysis of driver behavior by producing systems with an operational domain that is assessable by human drivers. We further conclude that the analysis of human-vehicle interaction can be systematically performed using a satisficing-based description of multiple mental model dynamics.

Through learning, adaptation, and optimization, operators have identified the most informative signals as input and have established performance criteria that produce behaviors that have proven to result in a satisficing tradeoff between various driver needs (e.g., expediency, safety, and comfort). Utilizing this knowledge in the design of automation is one step toward successful human-centered automation, but much work still needs to be done since modeling human behaviors in complex systems is still very challenging.

References

[1] L. Bainbridge, "Ironies of automation," *Automatica*, vol. 19, no. 6, pp. 775–779, 1983.
[2] E. R. Boer, "Tangent point oriented curve negotiation," *Proc. 1996 IEEE Intelligent Vehicles Symp.*, pp. 7–12, Sept. 19–20, 1996.
[3] E. R. Boer, E. C. Hildreth, and M. A. Goodrich, "A driver model of attention management and task scheduling: Satisficing decision making with dynamic mental models," presented at the Proc. 17th Euro. Annu. Conf. Human Decision Making Manual Control, Valenciennes, France, Dec. 14–16, 1998.
[4] ——, "Drivers in pursuit of perceptual and virtual targets," presented at the *IEEE Proc. Intelligent Vehicles Symp.*, Stuttgart, Germany, Oct. 28–30, 1998 pp. 291–296.
[5] E. R. Boer and M. Hoedemaeker, "Modeling driver behavior with different degrees of automation: A hierarchical decision framework of interacting mental models," presented at the 17th Euro. Annu. Conf. Human Decision Making Manual Control, Valenciennes, France, Dec. 14–16, 1998.
[6] J. J. Gibson and L. E. Crooks, "A theoretical field-analysis of automobile-driving," *Amer. J. Psychol.*, vol. 51, no. 3, 1938.
[7] M. A. Goodrich and E. R. Boer, "Maintaining driver autonomy with collision avoidance systems: A satisficing approach," Cambridge Basic Research, Nissan Research and Development, Inc., Cambridge, MA, Tech. Rep. TR-97-2, 1997.
[8] ——, "Semiotics and mental models: Modeling automobile driver behavior," presented at the Joint Conf. Science Technology Intelligent System, Gaithersburg, MD, Sept. 1998.
[9] ——, "Multiple mental models, automation strategies, and intelligent vehicle systems," in *Proc. IEEE/IEEJ/JSAI Int. Conf. Intelligent Transportation Systems*, Tokyo, Japan, Oct. 5–8, 1999, pp. 859–864.

[10] M. A. Goodrich, E. R. Boer, and H. Inoue, "Brake initiation and braking dynamics: A human-centered study of desired ACC characteristics," Cambridge Basic Research, Nissan Research Development, Inc., Cambridge, MA 02142, Tech. Rep. TR-98-5, 1998.

[11] ——, "A characterization of dynamic human braking behavior with implications for ACC design," in *IEEE/IEEJ/JSAI Int. Conf. Intelligent Transportation Systems*, Tokyo, Japan, Oct. 5–8, 1999, pp. 964–969.

[12] ——, "A model of human brake initiation behavior with implications for ACC design," in *Proc. IEEE/IEEJ/JSAI Int. Conf. Intelligent Transportation Systems*, Tokyo, Japan, Oct. 5–8, 1999, pp. 86–91.

[13] M. A. Goodrich, W. C. Stirling, and E. R. Boer, "Satisficing revisited," *Minds Mach.*, vol. 10, pp. 79–110, 2000.

[14] M. A. Goodrich, W. C. Stirling, and R. L. Frost, "A theory of satisficing decisions and control," *IEEE Trans. Syst., Man, Cybern. A*, vol. 28, pp. 763–779, Nov. 1998.

[15] ——, "Model predictive satisficing fuzzy logic control," *IEEE Trans. Fuzzy Syst.*, vol. 7, no. 3, June 1999.

[16] *Engineering Data Compendium: Human Perception and Performance*: Harry G. Armstrong Aerospace Med. Res. Lab., 1988, vol. II, sec. 7.3.

[17] P. N. Johnson-Laird, *The Computer and the Mind: An Introduction to Cognitive Science*. Cambridge, MA: Harvard Univ. Press, 1988.

[18] D. N. Lee, "A theory of visual control of braking based on information about time-to-collision," *Perception*, pp. 437–459, 1976.

[19] A. J. McKnight and B. B. Adams, "Driver Educations Task Analysis,", Springfield, MA, vol. 1, HumRRO Techn. Rep., 1970. Task Descriptions.

[20] J. A. Michon, Ed., "Generic Intelligent Driver Support: A Comprehensive Report on Gids," Taylor & Francis, New York, 1993.

[21] M. Minsky, *The Society of Mind*. New York: Simon and Schuster, 1986.

[22] N. Moray, "Designing for transportation safety in the light of perception, attention, and mental models," *Ergonomics*, vol. 33, no. 10/11, pp. 1201–1213, 1990.

[23] L. Nilsson, "Safety effects of adaptive cruise controls in critical traffic situations," in *Proc. Steps Forward: 2nd World Congress Intelligent Transport Systems*, Yokohama, Japan, Nov. 9–11, 1995, pp. 1254–1259.

[24] D. A. Norman, *The Design of Everyday Things*. New York: Doubleday, 1988.

[25] H. E. Pashler, *Engineering Psychology and Human Performance*. Cambridge, MA: MIT Press, 1997.

[26] J. Rasmussen, "Outlines of a hybrid model of the process plant operator," in *Monitoring Behavior and Supervisory Control*, T. B. Sheridan and G. Johannsen, Eds. New York: Plenum, 1976, pp. 371–383.

[27] J. Reason, "Cognitive aids in process environments: Prostheses or tools?," *Int. J. Man-Mach. Studies*, vol. 27, pp. 463–470, 1987.

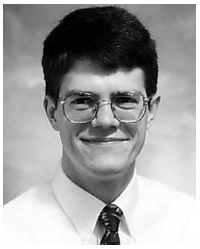
[28] G. N. Saridis, "Analytic formulation of the principle of increasing precision with decreasing intelligence for intelligent machines," *Automatica*, vol. 25, no. 3, pp. 461–467, 1989.

[29] N. Sarter, "Making coordination effortless and invisible: The exploration of automation management strategies and implementations," presented at the CBR Workshop Human Interaction Automated Systems, June 1, 1998.

[30] T. B. Sheridan, *Telerobotics, Automation, and Human Supervisory Control*. Cambridge, MA: MIT Press, 1992.

[31] C. D. Wickens and J. G. Hollands, *Engineering Psychology and Human Performance*, 3rd ed. Englewood Cliffs, NJ: Prentice-Hall, 2000.

[32] Y. Xiao, P. Milgram, and D. J. Doyle, "Planning behavior and its functional role in interactions with complex systems," *IEEE Trans. Syst., Man, Cybern. A*, vol. 27, pp. 313–324, May 1997.



Michael A. Goodrich received the Ph.D., M.S., and B.S. degrees (cum laude) in electrical and computer engineering from Brigham Young University, Provo, UT, in 1996, 1995, and 1992, respectively.

From 1996 to 1998, he was a Research Associate with Nissan Cambridge Basic Research, Nissan Research and Development, Inc., Cambridge, MA. Since 1998, he has been with the Computer Science Department at Brigham Young University where he is an Assistant Professor. His research interests include modeling and controlling intelligent systems, decision theory, multiple-agent coordination, human-centered engineering, fuzzy logic, and estimation theory.

Erwin R. Boer was born in The Netherlands in 1965. He received the M.S. degree in electrical engineering from Twente University of Technology, Twente, The Netherlands, in 1990, as well as the Ph.D. degree in electrical engineering from the University of Illinois, Chicago, in the area of manual control, in 1995.

After graduation, he worked at Nissan Cambridge Basic Research, Cambridge, MA, as a research scientist from 1995 to 2001, where he focused on driver modeling, driver support system development, and human navigation. In 2001, he worked for Wingcast as the head of the driving simulator lab, where he developed and evaluated multi-modal telematics interfaces. Currently, he has his own consulting firm and collaborates with many universities worldwide in the area of driver support systems. His main area of research interest is the development of driver support systems that synergistically augment the driving ecology.