Extending the Decision Field Theory to Model Operators' Reliance on Automation in Supervisory Control Situations

Ji Gao and John D. Lee

Abstract—**Appropriate trust in and reliance on automation are critical for safe and efficient system operation. This paper fills an important research gap by describing a quantitative model of trust in automation. We extend decision field theory (DFT) to describe the multiple sequential decisions that characterize reliance on automation in supervisory control situations. Extended DFT (EDFT) represents an iterated decision process and the evolution of operator preference for automatic and manual control. The EDFT model predicts trust and reliance, and describes the dynamic interaction between operator and automation in a closed-loop fashion: the products of earlier decisions can transform the nature of later events and decisions. The simulation results show that the EDFT model captures several consistent empirical findings, such as the inertia of trust and the nonlinear characteristics of trust and reliance. The model also demonstrates the effects of different types of automation on trust and reliance. It is possible to expand the EDFT model for multioperator multiautomation situations.**

Index Terms—**Automation, decision making, reliance on automation, supervisory control, trust in automation.**

I. Introduction

AUTOMATION can improve the efficiency and safety of complex and dangerous operating environments by reducing the physical or mental burden on human operators. However, operators do not always rely on automation appropriately, which can produce the opposite effect, causing inefficiency or undermining safety [1], [2]. Research examining the performance of people working with automated systems has received increasing attention as the problems associated with underreliance and overreliance on automation have resulted in several high-profile disasters [2]–[4].

Supervisory control systems represent situations in which automation and people work together to accomplish tasks [5]. Appropriate reliance on automation is an important factor affecting the performance of such systems, but only a few computational models have been developed to examine the factors influencing the decision to rely on automation [6]–[9]. The present study provides a quantitative model to describe the factors that affect reliance on different types of automation.

Manuscript received May 26, 2004; revised December 14, 2004 and April 15, 2005. This work was supported by the National Science Foundation under Grant No. 0117494. This paper was recommended by Associate Editor R. A. Hess.

The authors are with the Department of Mechanical and Industrial Engineering, University of Iowa, Iowa City, IA 52246 USA (e-mail: jgao@engineering.uiowa.edu; jdlee@engineering.uiowa.edu).

Digital Object Identifier 10.1109/TSMCA.2005.855783

Reliance on automation in a supervisory control situation represents decision making under risk and uncertainty. Under such conditions, it is not only cognitive factors but also emotional factors that influence decision making [10]–[12]. People react to the prospect of risk at two levels: they evaluate it cognitively, and react to it emotionally [10]. These two reactions have different determinants; specifically, cognitive evaluations of risk are sensitive to the objective features of the risky situation, such as probabilities of outcomes and assessments of outcome severity, whereas emotional reactions are sensitive to the vividness with which consequences can be imagined, or to recall of previous experience [10]. Trust is an important emotional factor influencing decision making as it relates to the decision to rely on automation [13].

Trust represents an affective response to the capability of the automation. Substantial evidence suggests that trust in automation is a meaningful and useful construct to understand the operators' reliance on automation. A review of different models of trust in automation shows that the current models suffer from a tradeoff between quantifying predictions and including a plausible psychological basis [14]. Conceptual and qualitative models describe the dynamics of trust between humans and automation [13], [15]–[17], and time series and regression models provide a quantitative description of trust, but are primarily statistical models that are fit to the data [1], [18]–[20]. Self-confidence in manual control capability, another critical factor in decision making [21], [22], can mediate the effect of trust on reliance [18]. A computational model of trust in automation and self-confidence can facilitate the understanding of the dynamics between trust and self-confidence, and how these influence reliance on automation. Another advantage of adopting a computational modeling approach is that it can generate testable predictions about human performance in different settings.

This paper provides the first step in developing a quantitative model of trust and reliance that is linked to rigorous psychological models of decision making. Specifically, we extend the decision field theory (DFT) [23] to characterize the operators' multiple sequential decisions in supervisory control situations in order to create a computational model that predicts trust, self-confidence, and reliance.

Such a model could complement the distinctions between types and levels of automation and guide automation design [4]. An important distinction between different types of automation is the information available regarding the capability of

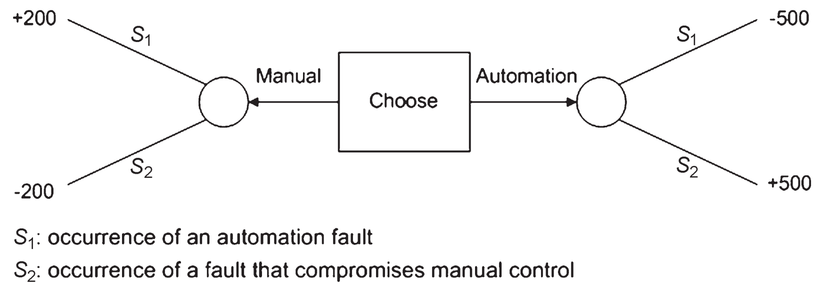


Fig. 1. Choosing to rely on automation (automation) or to intervene with manual control (manual) in a supervisory control situation.

the automation. In this study, two important factors regarding automation are addressed in the model. First, two types of automation are distinguished in terms of information availability regarding automation capability: mode-independent automation (MID) and mode-dependent automation (MD). The distinction describes whether the automation capability information is available when the automation is not engaged. Second, the transparency of the system interface is considered. System transparency describes how well the system interface reveals relevant information to help the operator estimate the state of the automation. This model helps quantify the conceptual distinctions regarding the types of automation to support design.

The paper is organized as follows. In Section II, the original DFT model is introduced using a simple example of supervisory control. To apply DFT to multiple sequential decision processes in supervisory control situations, extensions are needed. Section III presents the extensions of DFT that are needed to characterize dynamic decision making more appropriately. Section IV describes a controlled experiment and the results are compared to those generated by the extended DFT (EDFT) model. Several consistent empirical findings captured by the EDFT model are illustrated and the influence of each parameter on reliance is compared. In Section V, the effect of different types of automation on reliance is demonstrated. Section VI identifies important conceptual distinctions of the EDFT model and considers several other computational models that describe the human operator's reliance on automation. Conclusions and discussion are presented in Section VII.

II. DFT

DFT provides a rigorous mathematical framework for understanding the cognitive and motivational mechanisms that guide the human deliberation process in making decisions under uncertainty [23]. DFT was developed specifically to explain such behavior, and it differs from most mathematical approaches to decision making in that it is stochastic and dynamic rather than deterministic and static. This approach has been extended to explain the relationships between choice, selling prices, and certainty equivalents [24]. To extend these applications to multiple choice problems, a multialternative DFT was developed [25]. DFT has been applied successfully across a broad range of cognitive tasks, including sensory detection, perceptual discrimination, memory recognition, conceptual categorization, and preferential choice [26]. Comparison of DFT with other models has shown that DFT provides a better account of the basic empirical findings than approaches such as simple scalability models, standard random utility models, horse race random utility models, and elimination-by-aspects models [26].

The dynamic nature of DFT allows the model to describe the time course of decisions. As a consequence, DFT has accounted for several important experimental findings regarding decision making under uncertainty, such as the speed–accuracy tradeoff. The operator's choice between automatic and manual control in supervisory control situations can be considered both a preferential choice problem and a decision problem under uncertainty due to the complexity and variability of automation performance. Consequently, DFT offers an appropriate modeling approach to describe the dynamics of trust and the decision to adopt automatic or manual control.

The meaning of "dynamic" requires clarification in this context. In DFT and dynamic decision making, "dynamic" has two related, but subtly different, meanings. DFT is a dynamic approach in that it describes the time course of cognition (i.e., deliberation) preceding a decision. In contrast, dynamic decision making refers to the multiple and interdependent decisions made in an environment that changes autonomously and in response to a decision maker's actions [8], [27]. Each of these meanings of "dynamic" is critical in modeling reliance on automation. DFT is useful in describing a decision maker's interaction with a dynamically changing environment because it captures the continuous changes in the state of the decision maker. Such information is critical for capturing the autocorrelation of the decision maker's response to the dynamically changing systems. Specifically, DFT describes the time course of decisions in static situations—a critical first step in describing the time course of multiple interdependent decisions in dynamic situations. To date, there has been no application of DFT to this kind of dynamic decision making (e.g., multiple interdependent decisions in a dynamic system). Therefore, an extended version of the DFT model is needed to describe the decision of the operator to rely on automation in supervisory control situations.

We use a simple example in a supervisory control context to describe the basic idea of DFT. As shown in Fig. 1, the problem is to choose whether to rely on automation (A) or to intervene with manual control (M). One of two events may occur, noted by S_1 and S_2. S_1 denotes the occurrence of an automation fault and S_2 represents the incidence of a fault that compromises manual control. Let y_{ij} represent the payoff produced by taking action i ($i =$ A or M) when event S_j ($j = 1$ or 2) occurs and $u(y_{ij})$ represent the utility of the payoff y_{ij}.

For instance, when automatic control is chosen, the payoff will be -500 if a fault with the automation occurs and 500 if a fault compromises manual control. When manual control is chosen, the payoff will be 200 if a fault with automation occurs and -200 if a fault compromises manual control. The payoffs are arbitrary choices for the purpose of illustration, and normally distributed payoffs represent a more general case [23]. These payoffs are chosen to characterize the situation in which the benefit of appropriate use of automation and the cost of inappropriate use of automation are both higher than those of using manual control.

According to the Subjective Expected Utility theory [28], each expected payoff associated with each event is assigned a subjective probability weight $W(S_j)$. This weight reflects the amount of attention given to event S_j ($j = 1$ or 2). In a decision-making trial with multiple deliberation samples, for example, $W(S_j)$ can change from sample to sample because of attentional fluctuations. Consequently, the subjective expected payoff for each action also fluctuates from sample to sample, defining the valence of an action V_i ($i =$ A or M). The valences for actions i ($i =$ A or M) at sample n are defined as

$$V_{\mathrm{A}}(n) = W(S_1) \times u(-500) + W(S_2) \times u(+500) \quad (1a)$$

$$V_{\mathrm{M}}(n) = W(S_1) \times u(+200) + W(S_2) \times u(-200). \quad (1b)$$

The average subjective weight across samples is defined as $w(S_j) = E[W(S_j)]$. The average valences associated with average weights are

$$v_{\mathrm{A}} = w(S_1) \times u(-500) + w(S_2) \times u(+500) \quad (2a)$$

$$v_{\mathrm{M}} = w(S_1) \times u(+200) + w(S_2) \times u(-200). \quad (2b)$$

It follows that the valence difference can be decomposed into two parts: $V_{\mathrm{A}}(n) - V_{\mathrm{M}}(n) = [v_{\mathrm{A}} - v_{\mathrm{M}}] + \varepsilon(n)$, where the residual $\varepsilon(n)$ represents the change in valence difference produced by the moment-to-moment fluctuations in attention during deliberation. The mean valence difference is $d = v_{\mathrm{A}} - v_{\mathrm{M}}$.

In this study, the preference is defined as the operator's preference of A over M. The preference state at sample n, $P(n)$, is derived based on the accumulated valence difference. The new valence difference $V_{\mathrm{A}}(\mathrm{n}) - V_{\mathrm{M}}(\mathrm{n})$ is added to the previous preference state to produce a new preference state

$$\begin{aligned} P(n) &= P(n-1) + [V_{\mathrm{A}}(n) - V_{\mathrm{M}}(n)] \\ &= P(n-1) + [d + \varepsilon(n)]. \end{aligned} \quad (3)$$

The effect of valence difference on preference may vary depending on whether the recent samples (i.e., a recency effect) or earlier samples (i.e., a primacy effect) have greater impact. To take account of such serial position effects, the preference becomes

$$P(n) = (1-s) \times P(n-1) + [d + \varepsilon(n)] \quad (4)$$

where s is the growth–decay rate parameter, which determines the relative influence of the previous preference state and the new input on the current preference state. The preference update formula can be further extended when considering the effect of approach–avoidance conflict, which describes why avoidance–avoidance decisions take longer than approach–approach decisions [23]. This study uses a simplified version of the model in which the preference is updated based on (4) and the effect of approach–avoidance conflict is not considered.

The initial preference value $P(0)$ is given by z. The mean valence difference $d = v_{\mathrm{A}} - v_{\mathrm{M}}$ determines the direction of preference evolution. A decision is made once the preference goes beyond a threshold θ. Specifically, automatic control is adopted once the preference evolves beyond θ and manual control is adopted if the preference drops below $-\theta$.

Several similarities exist between trust in automation and the characteristics associated with the decision-making process described by DFT. First, trust and preference both evolve over time. Trust develops according to the reliability of the automation: faults cause trust to decline [1]. In DFT, preference is updated based on the decision maker's evaluation of the attributes of the alternatives. Second, trust and preference both have a stochastic characteristic in that they involve varying the attention to different events over time. Preference varies as the decision maker considers the different possible outcomes of choice alternatives. In a similar manner, trust varies over time as operators switch attention between different events [19]. Third, trust and preference both exhibit inertia, reflecting the process of information accumulation. The effect of faults on trust and the recovery afterwards is not instantaneous but occurs over time [1], [18]. This is consistent with one of the most important concepts of DFT regarding the preference evolution: the deliberation process involves an accumulation of information and the current preference depends on the previous values of preference. Fourth, trust and preference both have an initial bias. There are large individual differences in human–automation trust [18], [20], which contributes to the initial value of trust. Similarly, the initial value of preference is biased by prior knowledge and past experiences [23]. Self-confidence is another important factor influencing the operator's reliance on automation and has some similar characteristics to trust, such as inertia [18]. Self-confidence represents a state that changes from moment to moment and is distinguishable from self-efficacy, which is an enduring personality trait [29]. In all, the DFT description of preference is consistent with research that shows reliance depends on a dynamic interaction between trust in automation and the operator's self-confidence in manual control capability.

III. EDFT Applied to Supervisory Control

A. Description of EDFT Model

DFT has great potential to describe an operator's decision to rely on automation or to intervene with manual control in supervisory control situations. However, to apply DFT in describing such situations, two aspects of the original model must be extended.

The first concerns the growth model of the preference evolution in DFT. There are different ways to describe how preference grows over time. The simple random walk model [see (3)] represents unbounded growth as the mean preference increases linearly with time. This approach might be appropriate for situations in which the decisions are simple and the reaction times are short. However, a bounded growth of the preference might be a more realistic assumption to describe human decision-making behavior that extends over a relatively long time. There are different models for the bounded growth of preference. The DFT adopted a linear growth model [see (4)] that includes a growth-decay parameter (s) to provide a bounded growth of preference. In such cases, the mean preference converges to d/s.

An alternative linear growth model to (4) is described as

$$P(n) = (1-s) \times P(n-1) + s \times d + \varepsilon(n). \tag{5}$$

This is essentially an autoregressive model, and it guarantees that the preference will converge to d. It is a simple form of delta or error correction learning used in neural nets, which have been used to describe human decision making [30]. In the EDFT model, (5) is chosen to represent the supervisory control situation where the operator's preference is assumed to reflect the true difference between two alternatives given a sufficient amount of time.

The second aspect of DFT that needs to be extended is that DFT does not consider the effect of previous decisions in the context of multiple sequential decision processes. In a supervisory control situation, the operator's choice of automatic or manual control depends on the previous decision, and a new decision will influence the system state and the next decision.

In the EDFT model, sequential decision processes are linked by dynamically updating beliefs regarding the capability of automation or manual control based on previous experiences in order to guide the next decision. According to a framework developed in [31], beliefs represent the information base that determines attitudes (e.g., trust and self-confidence) and then attitudes determine intentions (e.g., preference) and consequently behaviors (e.g., reliance). The belief is updated as

$$B_{\mathrm{C}}(n) = \begin{cases} B_{\mathrm{C}}(n-1) + \frac{1}{b_1} \\ \quad \times \left(C(n-1) - B_{\mathrm{C}}(n-1)\right), & \text{if } \mathrm{INF_C} = 1 \\ B_{\mathrm{C}}(n-1) + \frac{1}{b_0} \\ \quad \times \left(B_{\mathrm{C_{ini}}} - B_{\mathrm{C}}(n-1)\right), & \text{if } \mathrm{INF_C} = 0 \end{cases} \tag{6}$$

where B_{C} represents the belief or estimation of the capability of automation (B_{CA}) or the operator's manual capability (B_{CM}). B_{Cini} is the initial value of B_{CA} or B_{CM}, and $B_{\mathrm{CA}}(0)$ and $B_{\mathrm{CM}}(0)$ represent the operator's initial belief in the automation and manual capability. C denotes the true capability of the automation (C_A) or manual control (C_{M}). The capability of automation (C_{A}) describes the reliability of the automation in terms of fault occurrence and general ability to accomplish the task under normal conditions. Similarly, the operator's manual capability (C_{M}) describes how well the operator can manually control the system in various situations. The difference between C_{A} and C_{M} corresponds to d in (5), that is, $d = C_{\mathrm{A}} - C_{\mathrm{M}}$. $\mathrm{INF_C}$ is a binary variable that represents whether the information of C is available to the operator. Namely, $\mathrm{INF_C} = 1$ corresponds to a situation in which information is available, and $\mathrm{INF_C} = 0$ corresponds to a situation in which it is not. The value b_1 represents the level of transparency of the system interface, describing how well information is conveyed to the operator when capability information is available; b_0, on the other hand, represents how strongly the belief depends on the operator's initial belief when capability information is not available.

$\mathrm{INF_C}$ plays an important role in defining different types of automation. Whether or not information regarding the capability of automation is available to the operator, independent of whether or not the automation is engaged, is a critical distinction. Information acquisition, information analysis, decision and action selection, and action implementation define four types of automation [4]. Different types of automation provide different degrees of feedback regarding the capability of automation. Information acquisition automation involves obtaining and integrating multiple sources of information whereas action implementation automation involves the execution of a course of action [4]. With information acquisition automation, the operator can observe the performance of the automation, even when it is not adopted. For example, in air traffic control (ATC), operators can always observe how the radar is working even when they do not accept the information it provides (the automation is not adopted). In contrast, with action implementation automation, the operator can only assess the capability of the automation when using it; the capability of the automatic sorting function of a photocopier can only be observed when it is relied upon. The $\mathrm{INF_C}$ parameter differentiates these types of automation.

In this study, two types of automation are defined in terms of whether having information available regarding automation capability depends on reliance or not: MID and MD automation. With MID automation, information about automation capability is always available no matter which control mode is used. With MD automation, information regarding capability is available only when automation is used and not available when the manual control is used. In terms of $\mathrm{INF_C}$, for MID automation, $\mathrm{INF_C} = 1$ is used to update B_{CA} independent of the reliance on automation. For MD automation, $\mathrm{INF_C} = 1$ is used to calculate B_{CA} when automation is used, and $\mathrm{INF_C} = 0$ is used when manual control is used. Similarly, the manual control can be MID or MD, depending on whether having information available regarding manual capability depends on the use of manual control. For MID manual control, $\mathrm{INF_C} = 1$ is used to update B_{CM} independent of the reliance on automation. For MD manual control, $\mathrm{INF_C} = 1$ is used to calculate B_{CM} when manual control is used and $\mathrm{INF_C} = 0$ is used when automation is used.

The evolution formula of preference shown in (5) is applied to trust and self-confidence, where B_{CA} and B_{CM} update trust and self-confidence as the new input

$$T(n) = (1-s) \times T(n-1) + s \times B_{\mathrm{CA}}(n) + \varepsilon(n) \tag{7}$$

$$\mathrm{SC}(n) = (1-s) \times \mathrm{SC}(n-1) + s \times B_{\mathrm{CM}}(n) + \varepsilon(n) \tag{8}$$

where T and SC correspond to trust and self-confidence. B_{CA} and B_{CM} are the input for the evolution of trust and self-confidence, representing the fact that automation and manual

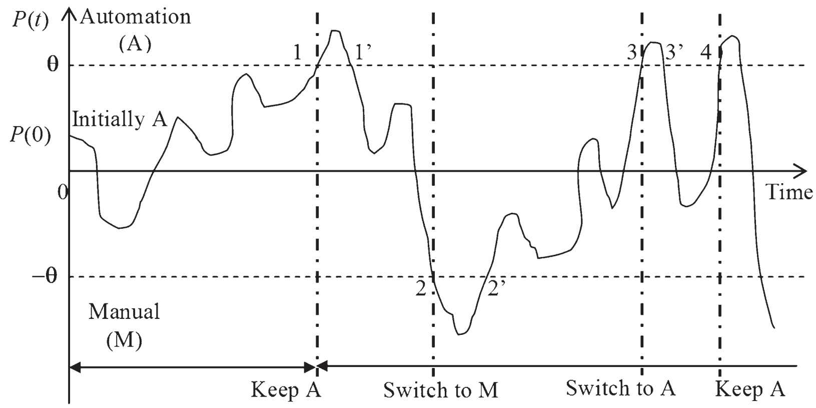


Fig. 2. Hypothetical trajectory showing how the threshold interacts with the changing level of preference.

control capabilities are the primary factors influencing the operator's decision to rely on automation or use manual control. The meaning of the growth–decay rate parameter s is the same as that in (4) and (5). s for T and SC are not necessarily the same, but are assumed to be the same for this simplified version of the model. ε is a random variable with zero mean and variance σ^2 that represents the uncertainty due to other factors besides the perceived capability that influences trust and self-confidence. The initial values of T and SC, $T(0)$ and $\mathrm{SC}(0)$, represent the initial bias towards automatic and manual control. The preference of A over M is defined as the difference between trust and self-confidence

$$P(n) = T(n) - \mathrm{SC}(n). \tag{9}$$

Equations (6) through (9) are used to estimate the preference in the model. To further illustrate the underlying relationship between the automation and manual control capabilities and the preference, the following equations are derived. Applying T and SC shown in (7) and (8) to (9), the preference updating formula becomes

$$P(n) = (1-s) \times P(n-1) + s \times [B_{\mathrm{CA}}(n) - B_{\mathrm{CM}}(n)] + \varepsilon_P(n) \tag{10}$$

where ε_p is a random variable with zero mean and variance σ_P^2. The noise term for preference in (10) is composed of the source of noise from both T and SC, so that σ_P^2 is considered the summation of the variance of each noise term ($\sigma_P^2 = 2\sigma^2$).

For purposes of illustration, the special case in which $\mathrm{INF_C} = 1$ and $b_1 = 1$ are used in (6) to update B_{C} is applied. It follows that $B_{\mathrm{CA}}(n) = C_{\mathrm{A}}(n-1)$, $B_{\mathrm{CM}}(n) = C_{\mathrm{M}}(n-1)$, and therefore $B_{\mathrm{CA}}(n) - B_{\mathrm{CM}}(n) = C_{\mathrm{A}}(n-1) - C_{\mathrm{M}}(n-1)$. As a result, (10) becomes

$$P(n) = (1-s) \times P(n-1) + s \times [C_{\mathrm{A}}(n-1) - C_{\mathrm{M}}(n-1)] + \varepsilon_P(n). \tag{11}$$

This is an equivalent formula of preference update to (5) with $d = C_{\mathrm{A}} - C_{\mathrm{M}}$. The decision to rely on automation or to use manual control occurs when the preference evolves beyond the threshold θ. Specifically, an operator who initially relies on automation switches to manual control once the preference evolves below $-\theta$; otherwise, the operator continues to rely on the automation. Fig. 2 illustrates this process. The operator keeps using automation and does not adopt manual control until reaching point 2, where the preference drops below $-\theta$, and switches back to automation at point 3, where the preference increases above θ.

In summary, the extension of DFT is important in that it characterizes multiple sequential decisions instead of the single decisions addressed by DFT. Based on the conceptual model [13], the EDFT model predicts trust and reliance and describes the dynamic interaction between operator and automation in a closed-loop fashion: the products of earlier decisions can transform the nature of later events and decisions. The operator's current decision to rely on automation determines what will be experienced next in terms of automation or manual performance, which subsequently influences the operator's belief, trust, self-confidence, and preference, in turn affecting the next decision on whether to rely on automation. Fig. 3 shows this closed-loop relationship between the context (C_{A} and C_{M}), information available (display), operator belief (B_{CA} and B_{CM}), attitude (T and SC), intention (P), and decision (reliance).

B. Model Parameterization

Fig. 3 shows six parameters of the EDFT model (given C_{A}, C_{M}, and $\mathrm{INF_C}$), namely, b_1, b_0, σ^2, s, θ, and z. The physical meaning of each parameter and its role in the closed-loop process are described in the context of a supervisory control situation.

b_1 represents the transparency of the system interface, describing how well the information is conveyed to the operator. Consequently, the level of b_1 determines how good the operator's estimation of the automation or manual control capability is compared to the true capability. b_1 is greater than or equal to 1. $b_1 = 1$ means the information is perfectly conveyed by the interface. In contrast, a larger value of b_1 indicates the information is poorly conveyed. Smaller values of b_1 should lead to more appropriate reliance on automation.

b_0 determines the relative influence of the operator's initial and most recent belief (belief at the previous time step) on current belief when the capability information is not available.

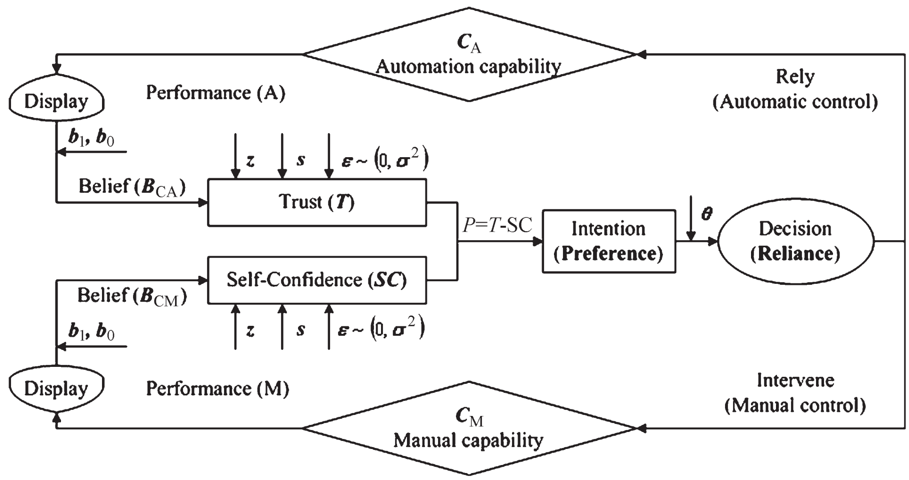


Fig. 3. Structure of the EDFT model of trust, self-confidence, and reliance on automation.

b_0 is greater than or equal to 1. $b_0 = 1$ represents that current belief depends completely on the initial belief, which does not change over time. A larger b_0 corresponds to situations where current belief depends less on the initial belief but more on the most recent belief.

σ^2 represents the variance of the noise term in (7) and (8). ε represents the uncertainty due to other factors besides the perceived capability influencing trust and self-confidence. The operator's attention to the automation might vary over time and lead to differences in trust and self-confidence. The uncertainties of trust and self-confidence account for the fluctuation of preference over time around the mean $d = C_{\mathrm{A}} - C_{\mathrm{M}}$.

s represents the growth–decay rate, describing how strongly the current state depends on the past state. This rate might be different for trust and self-confidence, but is assumed to be the same for the simplified version of the model. Mathematically, s defines the time constant that governs how quickly trust and self-confidence change. A larger value of s results in a greater weighting of new information and lower inertia of trust and self-confidence.

θ represents the inhibitory threshold, defining the minimum difference between trust and self-confidence needed to transition from manual to automatic control or vice versa. The lower the threshold, the more quickly and frequently operators alternate between the two types of control. With a higher threshold, a greater preference difference is required to make a decision. Thus, increasing θ increases the time required to make a decision. A higher θ represents situations where the decision is important, complex, or risky, and so demands that the decision maker deliberates thoroughly and thus possibly at greater length before making a decision [23].

z represents the initial value for the preference or the initial difference between trust and self-confidence, defined as $T(0) - \mathrm{SC}(0)$. At time zero, there is no observation of system behavior; therefore, individual bias constitutes the initial level of trust and self-confidence. A positive value of z represents a bias towards automation and a negative value a bias towards manual control. A small difference in the predisposition to trust may have a substantial effect if it influences an initial decision to engage automation [13]. In the context of supervisory control, these initial values represent the operator's bias regarding reliance on automation based on predisposition, previous working experience, and impressions of automation systems from others.

IV. Simulation Analyses

A. Model Validation

Data from an empirical study that examined the effect of augmenting a visual interface with sound on the operators' reliance on a semiautomatic process control system are used to validate this model [32]. The experiments in this study required operators to control a simulated orange juice pasteurization plant (Pasteurizer II), shown in Fig. 4 [33]. Realistic thermodynamic and heat transfer equations governed the dynamics of the process, and the simulation also incorporated some of the complexities of actual process control systems, such as time lags and feedback loops [1]. The simulated plant included provisions for both automatic and manual control. Operators could monitor and/or manipulate three control points: feedstock pump rate, steam pump rate, and heater setting. The operators chose either automatic control or manual control. Participants completed a series of 4-min trials. The system began in the automatic control mode, but operators could disengage the automation at any time. The operator's goal was to maximize production without burning or recycling juice by using automatic control, manual control, or any combination of the two.

The study compared the performance of operators in managing faults affecting automatic and manual control with and without the benefit of meaningful auditory feedback. It included four primary experimental conditions defined by two factors: noise and auto fault first, sound and auto fault first, noise and manual fault first, and sound and manual fault first. The first factor was a between-subjects variable, which defined the human–automation interface. In one interface condition, operators received uncorrelated noise, and in the other, continuous sound that was linked to the system state. The second factor was a within-subjects variable, defined as the reliability of the automatic and manual control. The order of the faults affecting

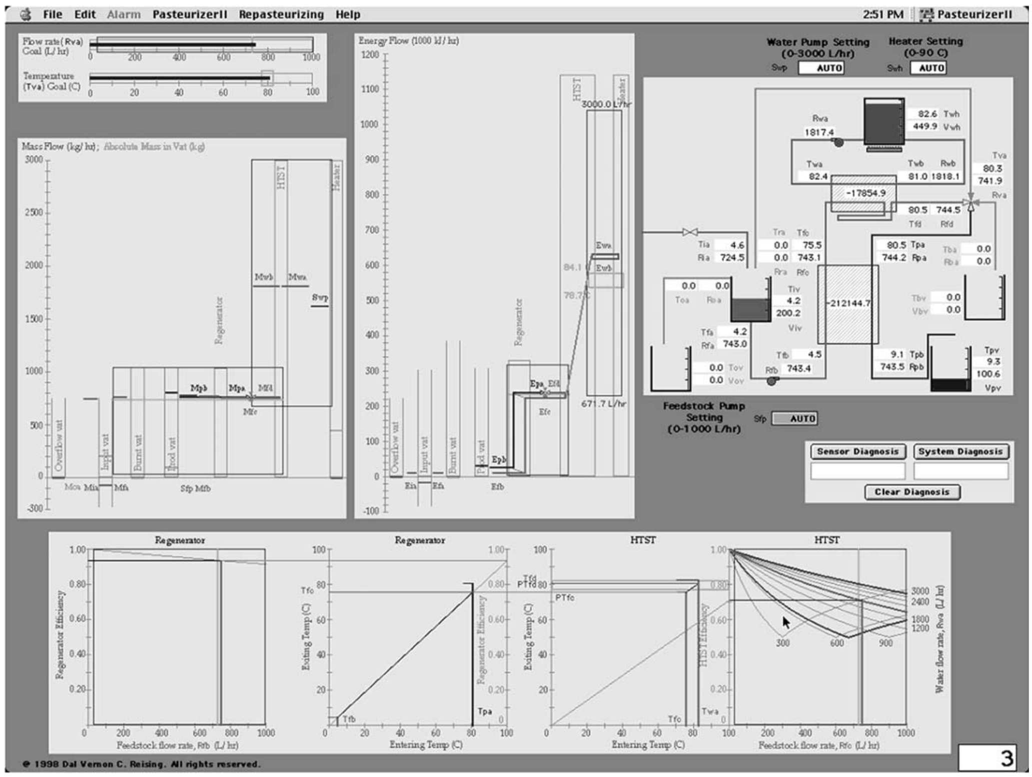


Fig. 4. Ecological interface for the Pasteurizer II microworld [33].

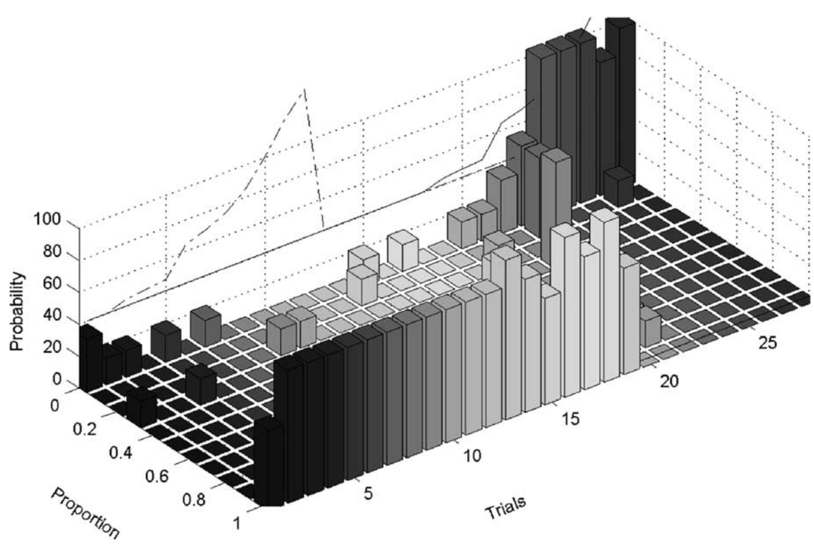


Fig. 5. Data for sound and manual fault first condition.

manual and automatic control was counterbalanced so that half the operators received unreliable automatic control followed by unreliable manual control. A total of 24 people participated. The data included 38 trials with each trial lasting 4 min. On the first day, participants received training and ran the plant for six trials, in which there were no faults. The first six trials of the second day also had no faults, but the next ten trials had faults in either automatic or manual control, with fault size increasing as the trials progressed. On the third day, the first six trials again had no faults, and the last ten trials had faults in either automatic or manual control, which again increased in size.

Empirical Data: The raw data were collected from the study [32] and are compared with the simulation results for each of four conditions. The data are displayed in the form of time-varying distribution of operators' reliance on automation. As an example, Fig. 5 shows the data for sound and manual fault first condition. A total of 28 sequential trials starting from the fifth trial of the second day are used for the validation. The proportion of reliance, with a scale of 0 to 1, represents the amount of time spent in automatic control during each trial (e.g., 0.2 represents 20% of time spent in automatic control during the trial). The vertical axis, with a scale of 0 to 100, corresponds to the percentage of operators who adopted each of the various

TABLE I
PARAMETER SETTINGS FOR MODEL VALIDATION

b_1	b_0	$\sigma^2(\Rightarrow\varepsilon)$	θ	s	z
2 or 10*	50	0.1 or 0.2*	$1.9\sqrt{2\sigma^2}$	0.5	0.3

* $b_1 = 2$ and $\sigma^2 = 0.1$ represent the sound condition and $b_1 = 10$ and $\sigma^2 = 0.2$ represent the noise condition.

levels of reliance for each trial. The solid and dashed curves shown on the vertical plane represent the occurrence and the magnitude of the faults affecting automatic and manual control (the curves are scaled in Fig. 5 to fit the graph). Specifically, the baseline corresponds to the normal condition and the peak corresponds to the most disruptive fault.

Parameter Selection: A total of 28 trials and 100 operators are simulated. Consistent with the experiment, the simulation starts with automatic control. Table I shows the settings of the six parameters used in the model validation. Four out of the six are essentially inherited from the original DFT model based on rules for mapping experimental factors into these parameters [23]. The specific value of each parameter is selected based on the rules as well as on a heuristic search using goodness of fit as the criterion, and so the resulting fit might not be optimal. The effects of six parameters on reliance are illustrated. The scenario of sound and manual fault first condition is used for the illustration and the corresponding data for this condition are shown in Fig. 5. When a particular parameter is manipulated to explore its effect, all the other parameters remain the same (see Table I).

b_1 represents the level of transparency of the system interface. The larger b_1, the less transparent the interface is, and therefore the less the operator can sense the automation capability. Simulations show that the operator can barely sense the gradual change of the capability when $b_1 \geq 100$. Fig. 6 compares the influences of $b_1 = 1$ and $b_1 = 100$ on reliance on automation. Because the pasteurizer system is not a perfectly transparent system, $b_1 = 1$ is not considered. The data indicate that participants can sense the automation capability better for sound condition than for noise condition. Therefore, a smaller b_1 is chosen for the sound condition than for the noise one. The degree of data fit suggests that $b_1 = 2$ and $b_1 = 10$ are appropriate values to represent the sound and noise conditions.

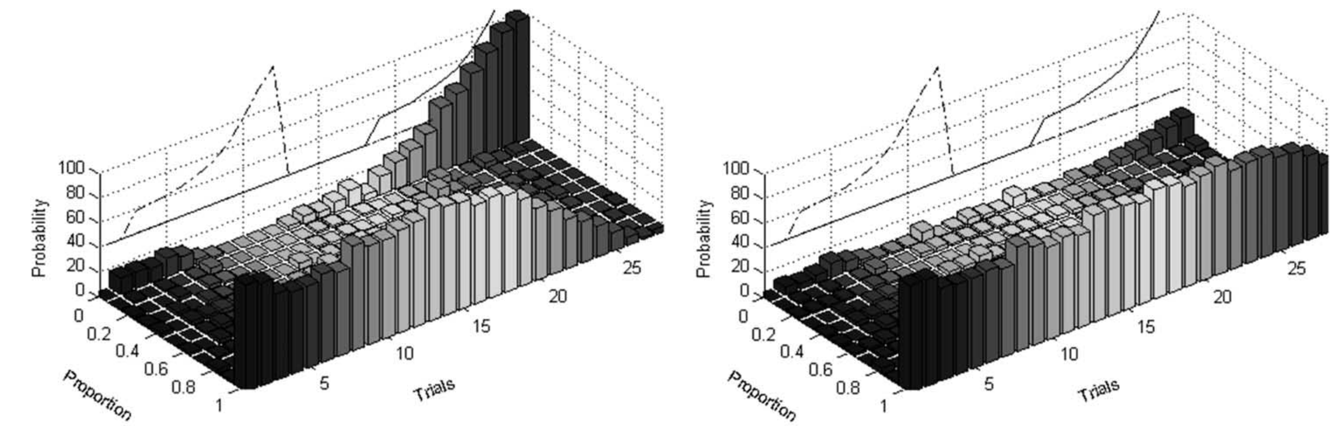
b_0 represents how strongly the current belief depends on the initial belief when capability information is not available. A medium value of b_0 (e.g., $b_0 = 50$) produces a slightly better fit compared to a low (e.g., $b_0 = 1$) or a high (e.g., $b_0 = 10^{10}$) value of b_0. Specifically, a 5% improvement of the fit for one condition and no improvement for the other three conditions are found. Although the small difference suggests that setting $b_0 = 1$ may greatly simplify the model, the parameter b_0 is included to generalize the model to other applications. For example, the influence of b_0 is expected to be more significant for situations where both automation and manual control are MD.

σ^2 accounts for the uncertainties of trust and self-confidence. The value of σ^2 depends on the average gain and loss for each choice based on the parameterization rules [23]. However, σ^2 is not derived this way because the payoff function is not easy to define for experiments that are used to verify the model. Equations (7) and (8) show that σ^2 is the variance of the noise term fluctuating around B_{CA} or B_{CM}; as a result, σ^2 is determined relative to the level of B_{CA} or B_{CM}. Fig. 7 compares the influence of different values of σ^2 on reliance. It shows that a smaller σ^2 reflects a situation where the operator is very certain about the choice—that is, a particular decision has a probability of almost 100%. In contrast, a larger σ^2 corresponds to a situation where the operator is uncertain. A noise condition has a larger degree of uncertainty so a larger σ^2 is used for the noise condition compared to the sound condition. A heuristic search suggests that $\sigma^2 = 0.1$ and $\sigma^2 = 0.2$ may represent the sound and the noise condition, respectively, to provide a good fit.

s reflects the inertia of trust and self-confidence. Fig. 8 compares the influence of two extreme values of s, 0 and 1, on reliance. $s = 0$ represents a situation where reliance is not influenced by an actual change in the automation or manual capability, but depends strictly on previous reliance and random variation. As shown in Fig. 8(a), after the first few trials, the operator either chooses A or M with equal probability, which reflects a random choice independent of the actual change of capability. In contrast, $s = 1$ represents a situation where reliance is strongly dependant on an actual change in capability, and therefore reliance is highly appropriate. A previous study of trust in automation with similar tasks found 0.5 to reflect the inertia of trust and found that value fitted the data well [1]. The same value 0.5 will be used here.

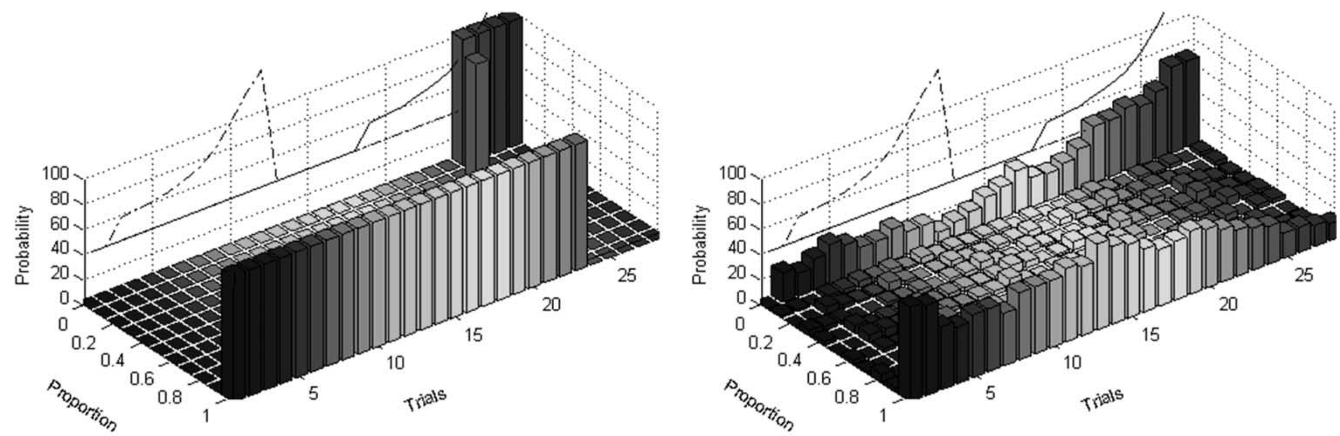
θ reflects the minimum difference between trust and self-confidence needed to transition from manual to automatic control or vice versa. Based on the rules of parameter determination [23], $\theta = f(L) \times \sigma_P$, where $f(L)$ is an increasing function of the time limit that constrains the decision making, L. In this study, time limit is not considered, and therefore a constant is used to replace $f(L)$. One interesting finding from the data is that participants spent either 0% or 100% of the time on either automatic or manual control within a trial with little in between. In the model, a threshold θ makes this phenomenon possible. Fig. 9 compares the influence of different values of θ on reliance. A low θ, such as $\theta = 0$, produces a frequent swap between automation and manual control during each trial, which is inconsistent with the data. A high θ, such as $\theta = 5\sqrt{2\sigma^2} = 0.7$, makes the initial decision permanent, which is also inconsistent with the data. Because θ is a threshold beyond which the preference evolves to determine a decision, an appropriate value of θ is between 0 and 2. Simulations suggest that $\theta = 1.9\sqrt{2\sigma^2} = 0.27$ represents the experimental data well.

z represents the initial preference and is equal to the initial difference between beliefs in automation and manual control capability. That is, $z = B_{\mathrm{CA}}(0) - B_{\mathrm{CM}}(0)$. In this model, a constant is used for z with a positive value corresponding to a bias towards automation and a negative value corresponding to a bias towards manual control. The data show that the operators tended to have a bias towards the automatic control in this task situation [18], [32]. Therefore, $z = 0.3$ is chosen for the simulation. On one hand, z influences the preference directly as shown in (10); on the other, z influences the preference



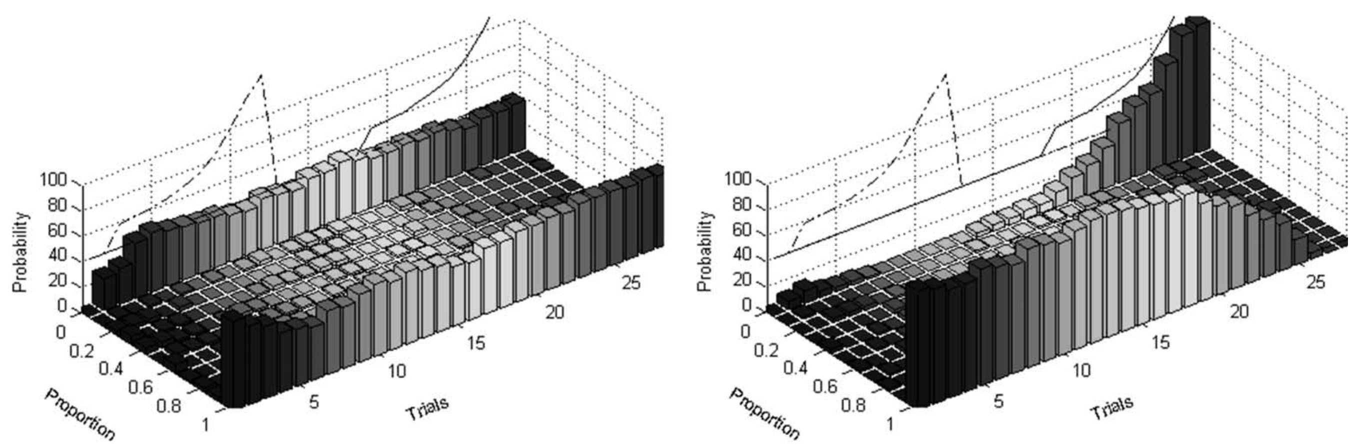
(a) (b)

Fig. 6. Influence of b_1 on reliance on automation. (a) $b_1 = 1$ (goodness of fit $= 80\%$); (b) $b_1 = 100$ (goodness of fit = 35%).



(a) (b)

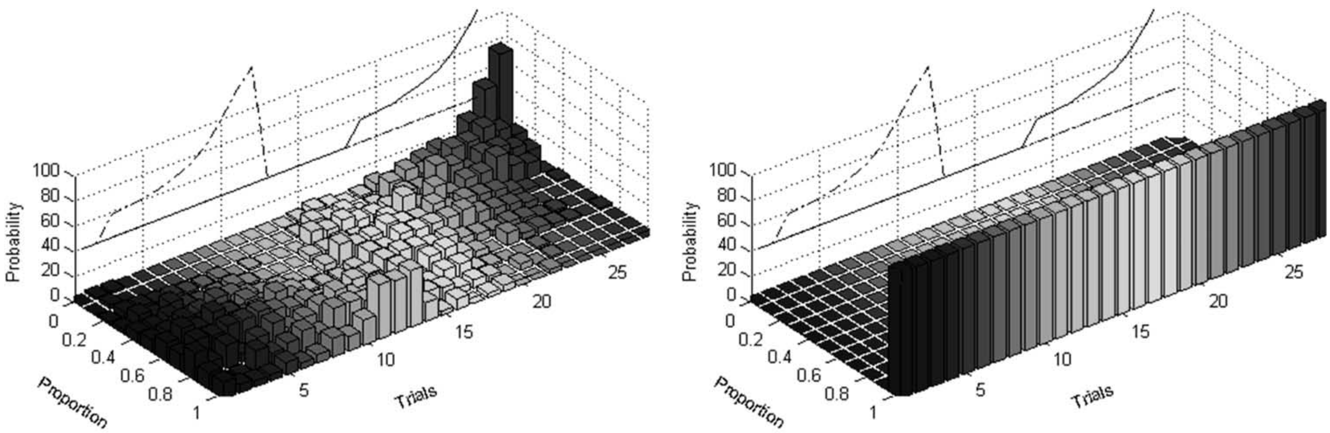
Fig. 7. Influence of σ^2 on reliance on automation. (a) $\sigma^2 = 0$ (goodness of fit $= 48\%$); (b) $\sigma^2 = 1$ (goodness of fit $= 65\%$).



(a) (b)

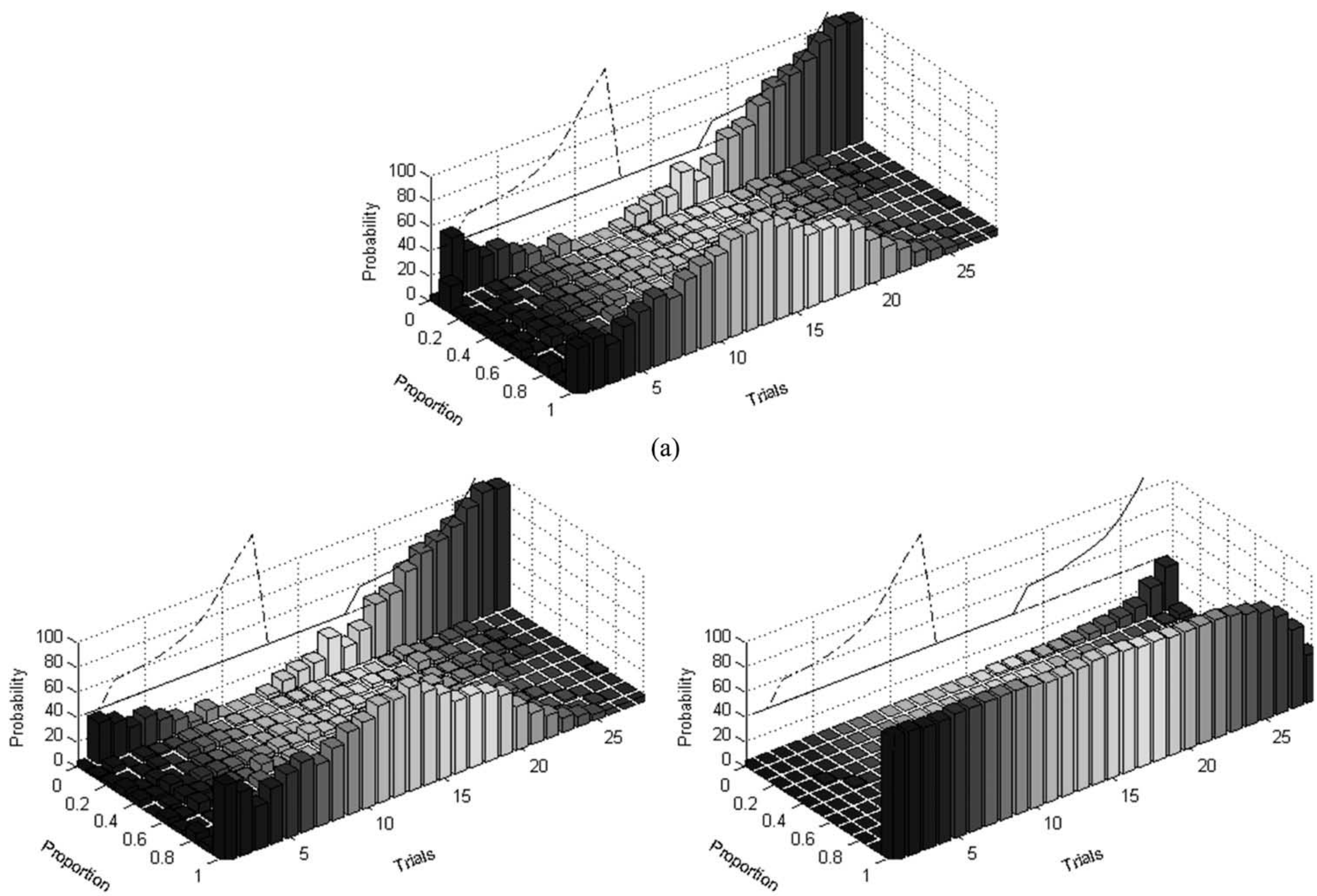
Fig. 8. Influence of s on reliance on automation. (a) $s = 0$ (goodness of fit $= 38\%$); (b) $s = 1$ (goodness of fit $= 77\%$).

indirectly because the corresponding $B_{\mathrm{C}}(0)$ influences the belief as in (6). Simulations show that the indirect influence of $B_{\mathrm{C}}(0)$ dominates the influence of z. A bias towards automation is associated with a lower $B_{\mathrm{CM}}(0)$ and a bias towards manual control is associated with a lower $B_{\mathrm{CA}}(0)$. Fig. 10 compares situations with a bias towards manual control (negative z), no bias (zero z), and a bias towards automation (positive z). The influence of a bias towards manual control is not significant because $B_{\mathrm{CA}}(0)$ influences reliance only for MD automation and the automation used in this study is



(a) (b)

Fig. 9. Influence of θ on reliance on automation. (a) $\theta = 0$ (goodness of fit $= 6\%$); (b) $\theta = 5\sqrt{2\sigma^2}$ (goodness of fit $= 4\%$).



(a) (b) (c)

Fig. 10. Influence of z on reliance on automation. (a) Bias towards manual control: $z = -0.8$ [i.e., $B_{\mathrm{CA}}(0) = 0.2$ and $B_{\mathrm{CM}}(0) = 1$] (goodness of fit $=$ 75%); (b) no bias: $z = 0$ [i.e., $B_{\mathrm{CA}}(0) = 1$ and $B_{\mathrm{CM}}(0) = 1$] (goodness of fit $= 77\%$); (c) bias towards automatic control: z $= 0.8$ [i.e., $B_{\mathrm{CA}}(0) = 1$ and $B_{\mathrm{CM}}(0) = 0.2$] (goodness of fit $= 29\%$).

considered as MID automation. In contrast, the influence of a bias towards automation is significant because $B_{\mathrm{CM}}(0)$ influences reliance only for MD manual control, and the manual control in this study is MD.

Model Input: The inputs of the model include variations in automation capability, in the operator's manual capability over time (C_{A} and C_{M}), and in the type of automation and manual control (MID or MD).

It is assumed that the occurrence of automation or manual faults completely determines the automation or manual capability. The experimental design detailed the timing and magnitude of the faults, which define changes in C_{A} and C_{M}.

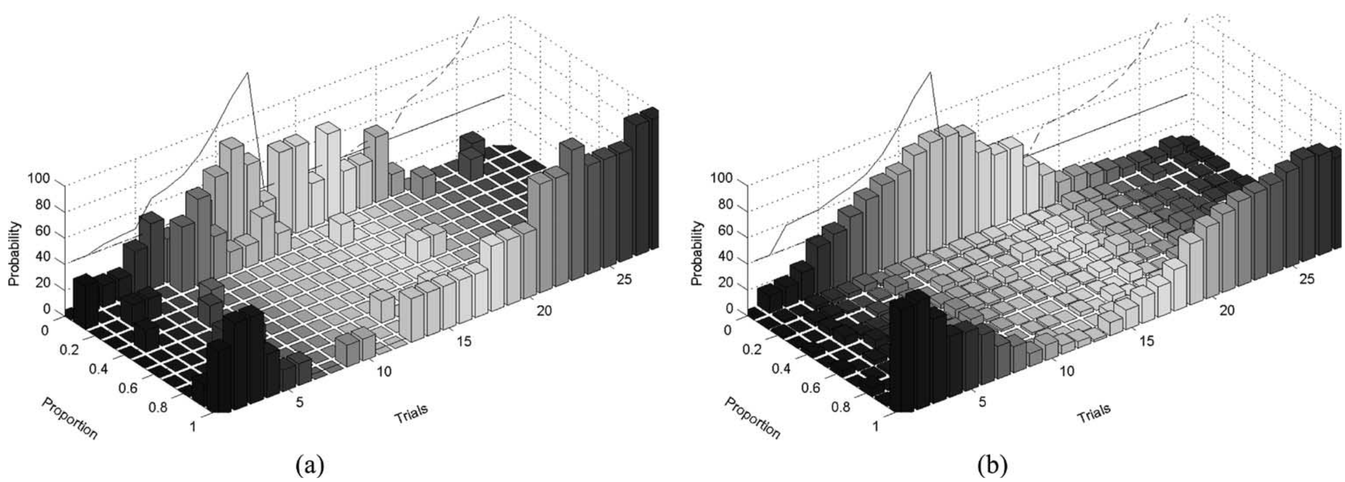


Fig. 11. Noise and auto fault first (goodness of fit $= 74\%$). (a) Data; (b) model.

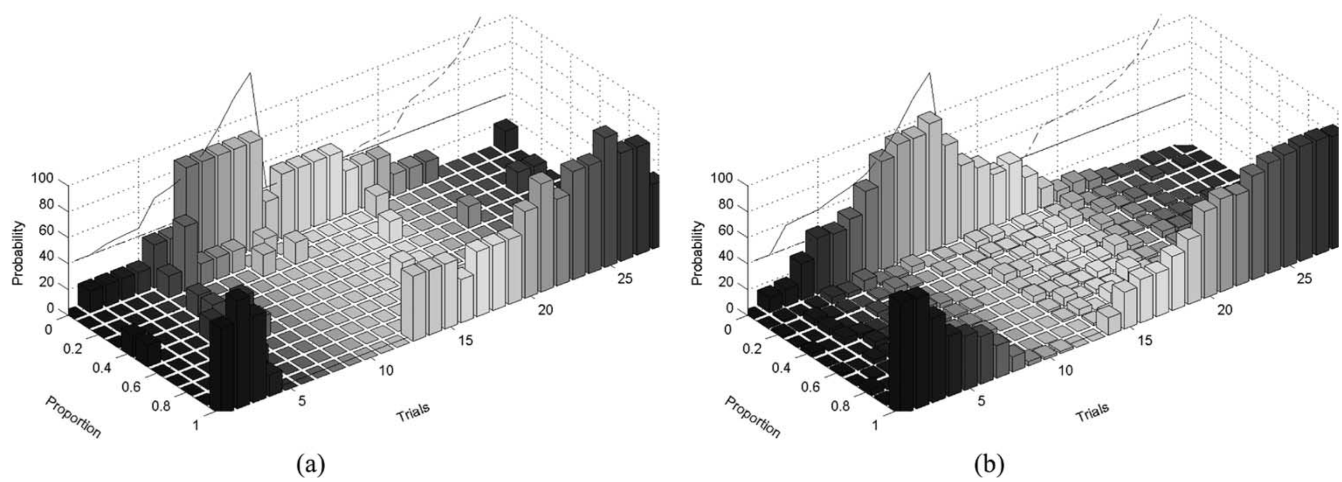


Fig. 12. Sound and auto fault first (goodness of fit $= 77\%$). (a) Data; (b) model.

The system operates under automatic control initially, but participants could freely choose either automatic or manual control at any time within each trial. Because each trial starts with automatic control and automation capability remains the same within the trial, the participant always had the opportunity to access the automation capability. Since this is the same as the situation in which an operator is able to access the automation capability even when the automation is not engaged, MID is used to describe the automation in this situation. That is, $\mathrm{INF_C} = 1$ is used to update B_{CA} independent of the control mode being used. In contrast, MD is used to describe the manual control condition because manual control performance can be observed only if the operator disengages the automation.

Model Output: The output of the model is the probability of using automatic control for a certain percentage of time during the trial. Consistent with the pattern seen in Figs. 11(a)–14(a), Figs. 11(b)–14(b) show that the distribution of operators' use of automation follows the occurrence of the fault and the corresponding fault size. For example, when the automation fault occurs (solid line), operators start to switch from automatic to manual control, and more operators move from automatic to manual control as the fault size increases.

Model Fit: The comparison between Figs. 11(a)–14(a) and Figs. 11(b)–14(b) demonstrates that the simulation captures the distribution of operators' reliance. Consistent with the empirical results, the simulations show that operator reliance on automation is jointly determined by the capability of automatic and manual control. The goodness of fit, as measured by the percent of variance accounted for by the model, shows that the model accounts for a large part of the variance of the data: 74%, 77%, 77%, and 79% for the four conditions shown.

B. *Effect of Inertia of Trust and Reliance*

Many studies of trust in automation have shown that trust demonstrates inertia [1], [9], [18]. The effect of faults on trust is not instantaneous: faults cause trust to decline over time. Likewise, the recovery of trust after faults is not instantaneous but occurs over time [1]. A time series analysis of the experimental data shows that trust has inertia, being dependent not only on the current size of faults and levels of performance but also on recent values of performance, fault size, and trust [1]. Reliance on automation likewise exhibits inertia [18]. The time series model of trust, self-confidence, and use of automation

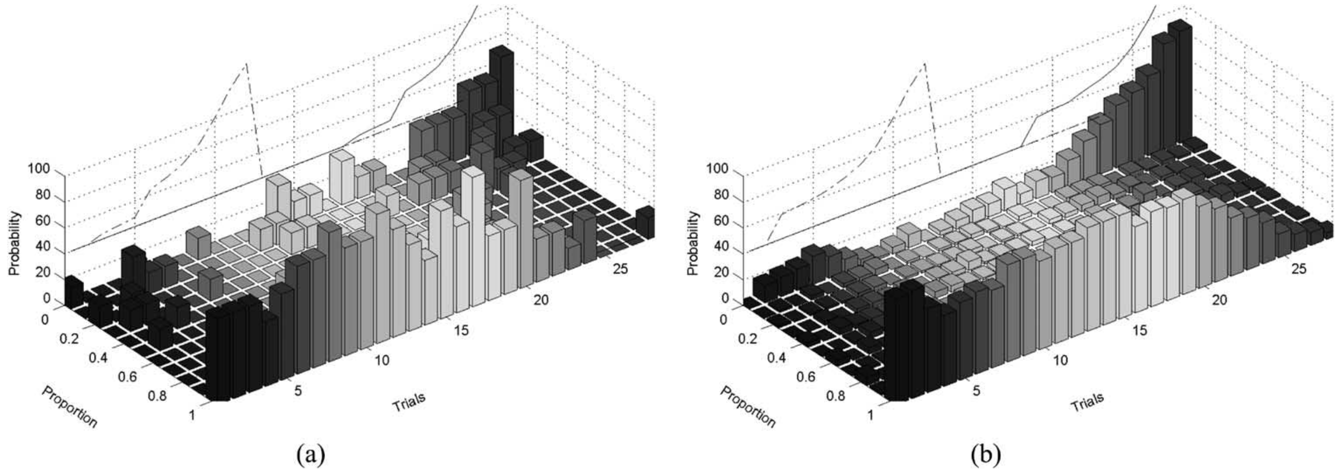


Fig. 13. Noise and manual fault first (goodness of fit $= 77\%$). (a) Data; (b) model.

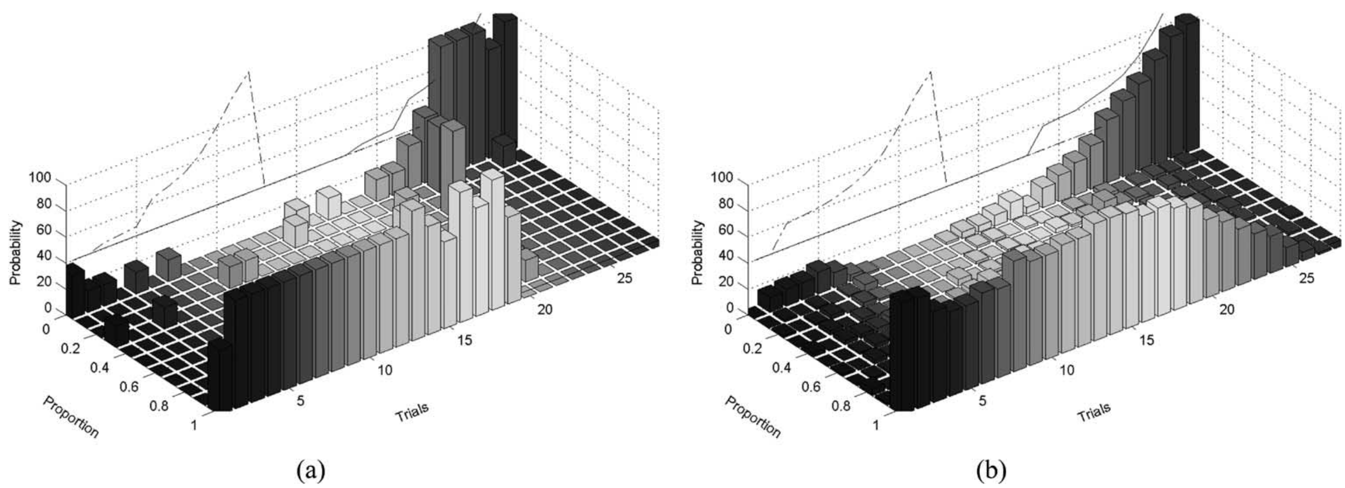


Fig. 14. Sound and manual fault first (goodness of fit $= 79\%$). (a) Data; (b) model.

indicates that the use of automatic control depends not only on the difference between trust and self-confidence but also on the previous use of the automatic control and the individual biases of operators [18]. The experimental data also show that trust is better calibrated under the sound condition compared to noise condition [32]. In terms of inertia, the data indicate that trust had less inertia when the automation fault occurred for the sound condition. Similarly, reliance also showed inertia, which was less for the sound condition [32].

Simulation: Fig. 15 shows the inertia effect for both data and simulation results for trust and self-confidence under the sound and manual fault first condition. The mean of data from six participants is used in Fig. 15(a). The sound condition defined by a smaller b_1 shows a smaller inertia effect of trust compared to the noise condition, which is consistent with the data. The data for sound condition show that self-confidence dropped slightly when manual faults occurred and increased slightly when the system returned to normal, but increased significantly when manual control was used during automation faults. This trend is well captured by the model, as shown in Fig. 15(b). The correlation coefficient between the model and the data is 0.90 for trust and 0.86 for self-confidence for this scenario, as shown in Fig. 15. For the noise and manual fault first scenario, the correlation coefficient is 0.89 for trust and 0.47 for self-confidence. Analysis of the data from individual participants for this scenario explains why the correlation coefficient for self-confidence is relatively low. The data show that one participant behaved differently from the others, with self-confidence declining drastically with the onset of the fault. This behavior accounts for much of the discrepancy between the model and the data. For the auto fault first scenarios, the correlation coefficient is 0.73 for trust and 0.82 for self-confidence (same for sound and noise conditions). Therefore, overall, the model accounts for the dynamics of trust and self-confidence well.

C. Nonlinear Relationship Between Trust, Self-Confidence, and Reliance

Previous studies have shown that the use of automatic control as a function of the difference between trust and self-confidence may not follow a linear relationship and that a logit function

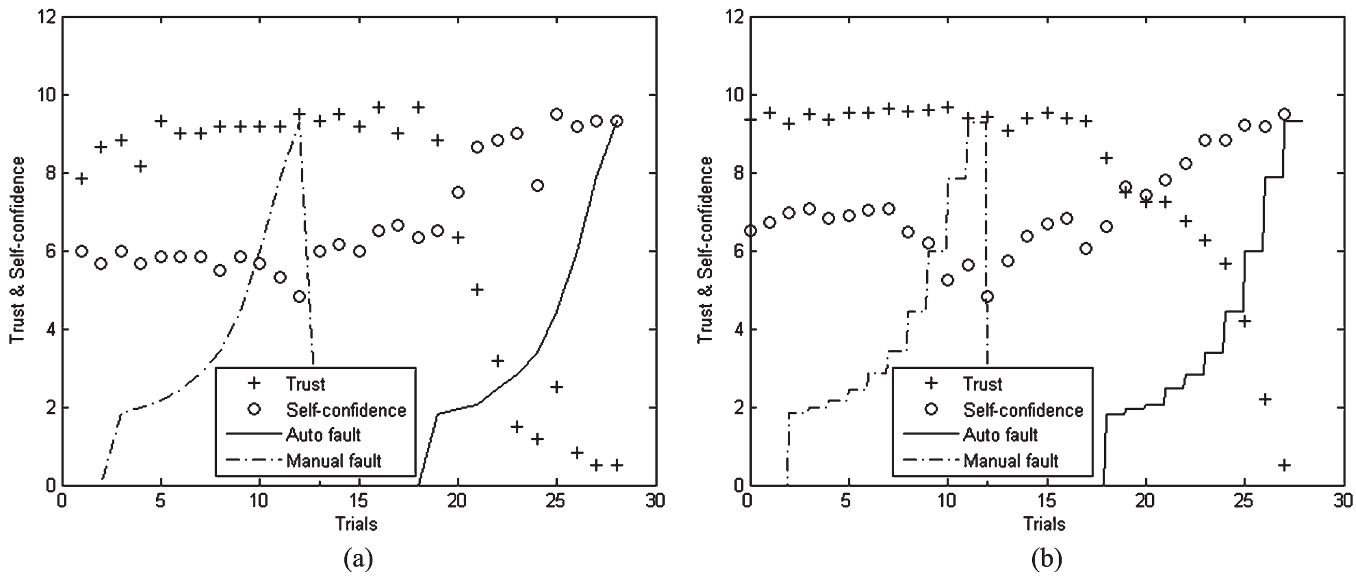


Fig. 15. Trust and self-confidence for sound and manual fault first condition. (a) Data; (b) model.

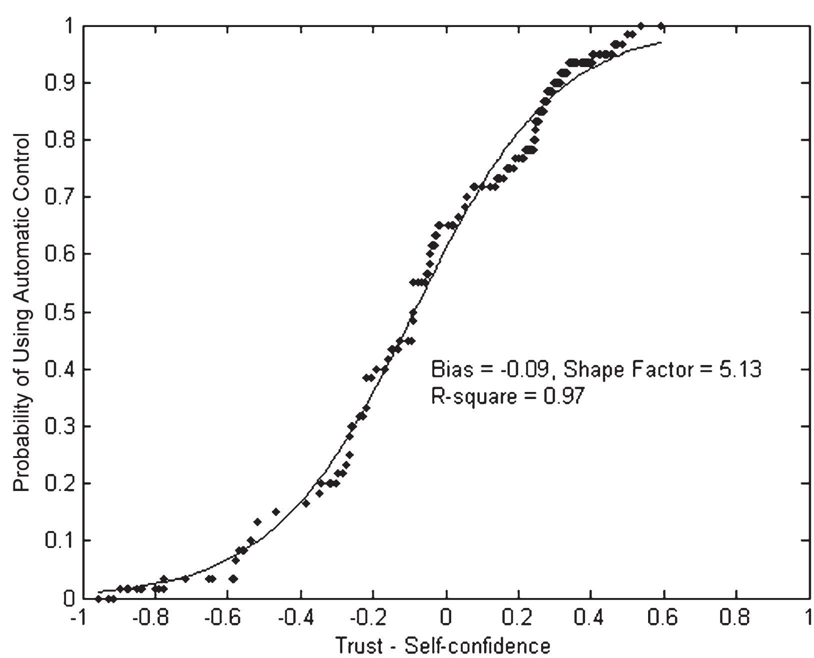


Fig. 16. Simulation results of the use of automation as a function of the difference between trust and self-confidence.

better represented the relationship [1], [18]. That is, complete use of automatic control is more likely adopted when trust is greater than self-confidence and complete manual control is more likely adopted when trust is less than self-confidence [18]. Operators seldom engaged the automation for only part of a trial. A logit function describes the nonlinear relationship between the use of automatic control and the difference between trust and self-confidence.

The strong tendency for complete automatic or complete manual control is evident in Figs. 11(a)–14(a). Fig. 16 shows that the simulation results also follow a logit function similar to previous empirical results ($R^2 = 0.97$). The logit function can be written as $y = 100/(1 + e^{-s(x-b)})$, where "$x$" is the difference between trust and self-confidence and "y" is the probability of using automatic control. Two parameters are required: "s" governs the shape of the function and "b" determines the bias. The bias parameter is slightly less than zero (-0.09), which suggests that for a given difference in trust and self-confidence, operators are more likely to use automatic than manual control. This result reflects the operator's bias towards the automatic control in the model. The shape factor of 5.13 suggests that the use of automatic control changes relatively quickly as a function of the increasing difference in trust and self-confidence. The good fit of the logit function is consistent with the general predictions of DFT [23]. Fig. 17 is also consistent with the substantial empirical results [18], [32],

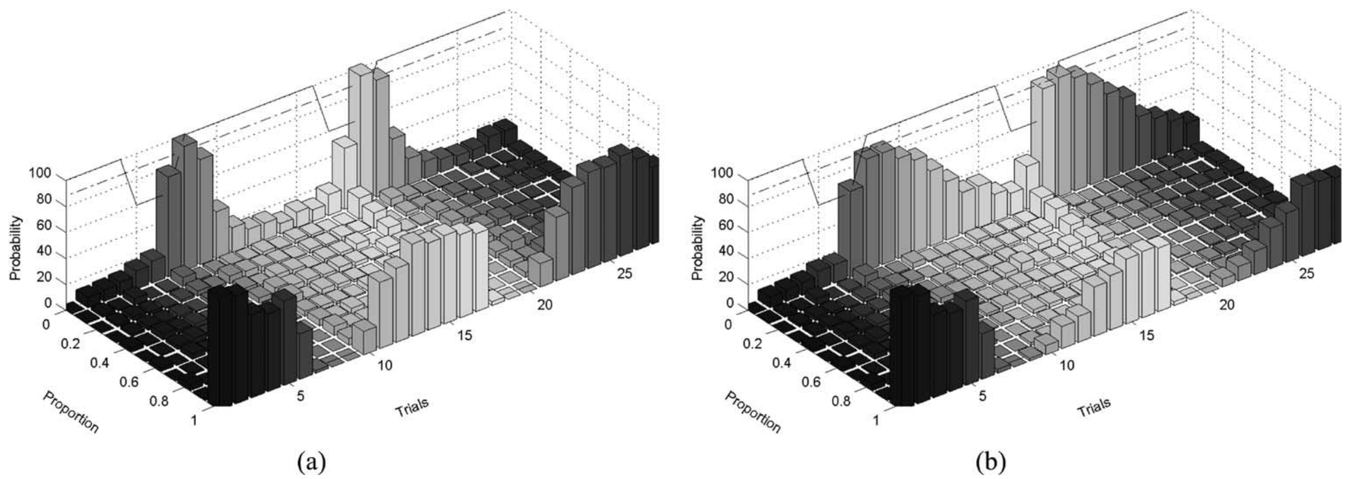


Fig. 17. Effect of different types of automation on reliance. (a) MID (information acquisition automation); (b) MD (action implementation automation).

which show a logit function describing the difference between trust, self-confidence, and reliance as a nonlinear relationship.

V. Further Experiments

The model can be used as a tool to guide the design of controlled experiments. One such application of the model is to show the effect of different types of automation on reliance, which is illustrated with the following analysis.

We have already discussed the distinction between MID and MD automation. Information acquisition automation represents MID and action implementation automation represents MD automation. These two types of automation may have very different effects on operators' trust and reliance, particularly in the context of faults.

Model Parameterization: The parameters are the same as those for the sound condition listed in Table I except that $b_1 = 1$ is used. Fig. 17 shows the fault pattern where C_{A} is represented by the solid line and C_{M} by the dashed line. C_{M} remains the same and C_{A} varies over trials. It represents a situation where automation outperforms manual control when no faults occur but underperforms manual control when faults occur. MID is used for the manual control. The MID and MD automation are distinguished by the belief-updating process. Specifically, for MID automation, $\mathrm{INF_C} = 1$ is used to update B_{CA} independent of the control mode being used. For MD automation, $\mathrm{INF_C} = 1$ is only used when automation is adopted, and $\mathrm{INF_C} = 0$ is used when manual control is adopted.

Simulation: Fig. 17 compares the distribution of reliance on automation for MID and MD types of automation. The figure shows that there is no difference in reliance until automation returns to normal from faults. Once this occurs and automation performs better than manual control, operators are still more likely to keep using manual control because of the inertia of trust and reliance. However, with MID automation, they would have chance to know the true capability of automation and to rely on it again. In contrast, with MD automation, they do not have access to the automation capability when manual control is used. Operators detect changes in automation capability more quickly with MID automation and are therefore more likely to rely on the automation appropriately. These results can help explain the conflicting findings of several studies in which trust in and reliance on automation recovers quickly in some situations and not others. For example, trust failed to recover fully when the reliability of automation that controls a pump improved [1] [Fig. 17(b)]. In contrast, other studies found that participants did not delay using automation after it recovered from failures [34], [35] [Fig. 17(a)]. The type of automation may be one reason for the discrepancy.

VI. Model Comparison

Relatively few computational models exist that predict operator reliance on automation, which is characterized by multiple sequential decisions in a dynamic environment that changes autonomously and in response to a decision maker's action. EDFT makes a unique contribution to modeling reliance on automation.

Gonzalez, Lerch, and Lebiere conducted a study that investigated the recognition process (i.e., the ability to discriminate among familiar classes of objects) in dynamic decision making [36]. The microworld-based decision-making task required participants to activate and deactivate the pumps for a water purification plant with 22 tanks, each with two pumps, in order to distribute all the water in the allotted amount of time. They examined how individuals' recognition ability changed as the similarity of decisions they made changed. Although this study characterizes multiple and interdependent real-time decisions in an autonomously changing environment, the environment is highly consistent in terms of dynamics (i.e., the time patterns with which exogenous events occurred are consistent). In contrast, the operator's supervisory control task considered in the EDFT model characterizes multiple sequential decisions in an uncertain environment, where automation reliability changes unexpectedly. That is, in terms of environmental consistency, their study describes decisions made repeatedly when confronted with consistent dynamic situations whereas the EDFT model describes decisions in a dynamic and uncertain environment. In terms of the five learning mechanisms crucial to skill development in dynamic decision making [37], they focused on instance-based knowledge and recognition-based retrieval whereas the EDFT model considers the influence of feedback on sequential decisions.

Kirlik, Miller, and Jagacinski developed a computational model that integrated a representation of the external environment with a representation of skilled human decision making to describe the human–automation interaction [6]. They modeled how both environmental and cognitive factors determine the decision. In the model, affordance values represent the degree to which the environmental structure afforded the action and priority values represent the degree to which the action was desirable in terms of task goals. These two factors were combined to calculate the appropriateness value for taking an action, which determines the decision. Although the model was used to describe the supervisory control task in a dynamic and uncertain environment, the interdependency between the previous action and the next decision was not considered. However, an important characteristic of the EDFT model is that it considers how the consequences of the previous action influence the next decision, thus linking the multiple sequential decisions in a dynamic way.

There are also several computation models that use "black-box" approaches, such as neural networks, which focus on the inputs and outputs rather than on the processes linking the two. Such approaches first develop the relationship between inputs and outputs by training the model with existing data, and then predict the output given new input data using the well-trained model. For example, Gibson, Fichman, and Plaut developed a computational model using a neural network to model the participants' pattern of performance in training, control, prediction, transfer, and level of performance [8]. The task required that the participants manipulate an input to a hypothetical sugar factory to achieve a particular production goal over time—in other words, they faced a dynamic decision problem. Farrell and Lewandowsky developed a connectionist model of operators' complacency, which adopted a neural network approach [9]. In contrast to these machine learning approaches, the EDFT model is more psychologically oriented in that it not only predicts output but also describes the cognitive process. For instance, besides the prediction of reliance on automation, the dynamics of trust and self-confidence are also predicted.

An analysis conducted by Sheridan and Parasuraman shows that calculating the expected value of alternative choices, such as automatic and manual control, can provide a rigorous means of selecting the best reliance option [7]. In such an analysis, an analytical criterion is developed, based on standard statistical decision theory, to guide the allocation of a simple failure detection task to the human operator or to the automation. This decision analysis suggests the best solution among options from an economic perspective, providing guidance with regard to better use of automated systems. This normative solution does not reflect human decision-making behavior when facing such a choice problem, but these normative models can help guide the design. The EDFT model describes the cognitive process and the human operator's decision making with respect to reliance on automation.

In summary, two characteristics distinguish the EDFT model from most of the other computational models that describe decision making in the area of human–automation interaction. First, it accounts for the influence of the consequences of previous actions on the next decision for multiple sequential decisions in a dynamic uncertain environment. Second, it accounts for the cognitive process that may underlie these decisions.

VII. Conclusion and Discussion

Although many researchers have investigated the issue of trust in and reliance on automation in recent years, relatively few quantitative models of trust and reliance have been developed. The original decision field theory (DFT) model offers a promising basis for a quantitative model of trust and reliance. DFT is extended to describe multiple sequential decisions in the supervisory control situation. The extended DFT (EDFT) represents the iterated decision process and the operator's preference evolution more realistically than the original DFT model for supervisory control tasks. The unique contribution of the EDFT model is that it is based on psychological principles, characterizes the dynamic interaction between human and automation, and makes quantitative predictions of operators' reliance on automation. The model depicts the dynamic interaction between operator and automation in a closed-loop fashion that describes the relation between the operator, the state of the automation, and the interface through which the operator receives information about the capability of automatic and manual control. The EDFT model replicates several empirical results including the inertia of trust and the nonlinear relationship between trust, self-confidence, and reliance. Also, the model provides a guide for further experiments to address the effects of different types of automation.

The EDFT model provides an excellent fit to the experimental data, but this is not a sufficient test of its contribution [38]. The number of free parameters in the model must be justified. There are two main reasons why six parameters are used in the EDFT model. First, each parameter describes a specific cognitive phenomenon or process. Four of the six are inherited from the original DFT model. The two new parameters are important to the extension of the DFT model in that they describe how the consequence of a previous action influences the next decision in the sequential multiple decision processes, which the original DFT model does not consider. Second, not all the parameters are independent. There are only five free parameters. Specifically, the threshold θ is a function of another parameter σ^2. Therefore, to predict the decision for a sequential multiple decision problem as well as to describe the dynamics of not only reliance but also trust and self-confidence, a model with five free parameters is relatively parsimonious.

The limitations of the EDFT model should also be noted. First, we assume that the primary inputs of the model—automation and operator capabilities (C_{A} and C_{M})—are available. However, the concept of "automation capability" has not been standardized in the human factors literature, and such a simplistic one-dimensional description is unlikely to be easily estimated, which is especially true for complex systems. The reality is often a complex gradation between the two. In this study, we used the occurrence of faults (reliability) to represent the capability, which might not be appropriate for situations where other factors, such as environmental perturbations, significantly influence capability. More generally, the simplistic description of "automatic" and "manual" control does not apply to many systems. Second, the level of fit for the model validation is good but not perfect. It is commonly acknowledged that a good fit is not sufficient because it reveals nothing about the flexibility of the theory and the variability of the data [38]. For example, to test how sufficient the model is, a large set of data from different task settings is needed. Therefore, further validation with a greater range of experimental data would be useful.

The EDFT model described in this paper addresses the decision-making process for single operator single automation in a supervisory control system. The EDFT model could be expanded to fit multioperator multiautomation situations in order to investigate the effects on trust and reliance of sharing the automation information between operators [39]. These effects are of particular importance from the perspective of distributed cognition and team decision making [40].

Acknowledgment

The authors thank D. V. Reising for supplying the Pasteurizer II microworld. The authors also acknowledge the constructive advice of three reviewers and the editor.

References

[1] J. D. Lee and N. Moray, "Trust, control strategies, and allocation of function in human–machine systems," *Ergonomics*, vol. 35, no. 10, pp. 1243–1270, 1992.

[2] R. Parasuraman and M. Mouloua, "Automation and human performance. Theory and applications," in *Human Factors in Transportation*, B. Kantowitz, Ed. Mahwah, NJ: Lawrence Erlbaum Associates, 1996, p. 514.

[3] R. Parasuraman and V. Riley, "Humans and automation: Use, misuse, disuse, abuse," *Hum. Factors*, vol. 39, no. 2, pp. 230–253, 1997.

[4] R. Parasuraman, T. B. Sheridan, and C. D. Wickens, "A model for types and levels of human interaction with automation," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 30, no. 3, pp. 286–297, May 2000.

[5] T. B. Sheridan, *Humans and Automation*. New York: Wiley, 2002.

[6] A. Kirlik, R. A. Miller, and R. J. Jagacinski, "Supervisory control in a dynamic and uncertain environment: A process model of skilled human–environment interaction," *IEEE Trans. Syst., Man, Cybern.*, vol. 23, no. 4, pp. 929–952, Jul.–Aug. 1993.

[7] T. B. Sheridan and R. Parasuraman, "Human versus automation in responding to failures: An expected-value analysis," *Hum. Factors*, vol. 42, no. 3, pp. 403–407, 2000.

[8] F. P. Gibson, M. Fichman, and D. C. Plaut, "Learning in dynamic decision tasks: Computational model and empirical evidence," *Org. Behav. Human Decis. Process.*, vol. 71, no. 1, pp. 1–35, 1997.

[9] S. Farrell and S. Lewandowsky, "A connectionist model of complacency and adaptive recovery under automation," *J. Exper. Psychol., Learn., Mem., Cogn.*, vol. 26, no. 2, pp. 395–410, 2000.

[10] G. F. Loewenstein, E. U. Weber, C. K. Hsee, and N. Welch, "Risk as feelings," *Psychol. Bull.*, vol. 127, no. 2, pp. 267–286, 2001.

[11] A. Bechara, H. Damasio, D. Tranel, and A. R. Damasio, "Deciding advantageously before knowing the advantageous strategy," *Science*, vol. 275, no. 5304, pp. 1293–1295, 1997.

[12] R. J. Dolan, "Emotion, cognition, and behavior," *Science*, vol. 298, no. 5596, pp. 1191–1194, 2002.

[13] J. D. Lee and K. A. See, "Trust in technology: Designing for appropriate reliance," *Hum. Factors*, vol. 46, no. 1, pp. 50–80, 2004.

[14] N. Moray and T. Inagaki, "Laboratory studies of trust between humans and machines in automated systems," *Trans. Inst. Meas. Control*, vol. 21, no. 4/5, pp. 203–211, 1999.

[15] B. M. Muir, "Trust in automation 1: Theoretical issues in the study of trust and human intervention in automated systems," *Ergonomics*, vol. 37, no. 11, pp. 1905–1922, 1994.

[16] V. Riley, "A theory of operator reliance on automation," in *Human Performance in Automated Systems: Recent Research and Trends*, M. Mouloua and R. Parasuraman, Eds. Hillsdale, NJ: Lawrence Erlbaum, 1994, pp. 8–14.

[17] M. S. Cohen, R. Parasuraman, and J. Freeman, "Trust in decision aids: A model and its training implications," in *Proc. Command and Control Research and Technology Symp.*, Monterey, CA, 1998, pp. 1–37.

[18] J. D. Lee and N. Moray, "Trust, self-confidence, and operators' adaptation to automation," *Int. J. Human-Comput. Stud.*, vol. 40, no. 1, pp. 153–184, 1994.

[19] B. M. Muir and N. Moray, "Trust in automation 2: Experimental studies of trust and human intervention in a process control simulation," *Ergonomics*, vol. 39, no. 3, pp. 429–460, 1996.

[20] N. Moray, T. Inagaki, and M. Itoh, "Adaptive automation, trust, and self-confidence in fault management of time-critical tasks," *J. Exp. Psychol., Appl.*, vol. 6, no. 1, pp. 44–58, 2000.

[21] A. Bandura, "Self-efficacy mechanism in human agency," *Amer. Psychol.*, vol. 37, no. 2, pp. 122–147, 1982.

[22] M. E. Gist and T. R. Mitchell, "Self-efficacy—A theoretical-analysis of its determinants and malleability," *Acad. Manage. Rev.*, vol. 17, no. 2, pp. 183–211, 1992.

[23] J. R. Busemeyer and J. T. Townsend, "Decision field theory: A dynamic cognitive approach to decision making in an uncertain environment," *Psychol. Rev.*, vol. 100, no. 3, pp. 432–459, 1993.

[24] J. T. Townsend and J. R. Busemeyer, "Dynamic representation of decision-making," in *Mind as Motion*, R. F. Port and T. V. Gelder, Eds. Cambridge, MA: MIT Press, 1995, pp. 101–120.

[25] R. Roe, J. R. Busemeyer, and J. T. Townsend, "Multialternative decision field theory: A dynamic connectionist model of decision-making," *Psychol. Rev.*, vol. 108, no. 2, pp. 370–392, 2001.

[26] J. R. Busemeyer and A. Diederich, "Survey of decision field theory," *Math. Soc. Sci.*, vol. 43, no. 3, pp. 345–370, 2002.

[27] B. Brehmer, "Strategies in real-time, dynamic decision making," in *Insights in Decision Making*, R. M. Hogarth, Ed. Chicago, IL: Univ. Chicago Press, 1990, pp. 262–279.

[28] W. Edwards, "Subjective probabilities inferred from decisions," *Psychol. Rev.*, vol. 69, no. 2, pp. 109–135, 1962.

[29] A. Bandura and D. Cervone, "Self-evaluative and self-efficacy mechanisms governing the motivational effects of goal systems," *J. Pers. Soc. Psychol.*, vol. 45, no. 5, pp. 1017–1028, 1983.

[30] J. R. Busemeyer and I. J. Myung, "An adaptive approach to human decision making: Learning theory, decision theory, and human performance," *J. Exp. Psychol., Gen.*, vol. 121, no. 2, pp. 177–194, 1992.

[31] M. Fishbein and I. Ajzen, *Belief, Attitude, Intention, and Behavior*. Reading, MA: Addison-Wesley, 1975.

[32] K. A. See, *The Effect of Sound on Performance, Reliance, Decision Making, and Trust in Semi-Automatic Process Control Systems*. Iowa City, IA: Dept. Mech. Ind. Eng., Univ. Iowa Press, 2002.

[33] D. V. C. Reising and P. M. Sanderson, "Ecological interface design for pasteurizer II: A process description of semantic mapping," *Hum. Factors*, vol. 44, no. 2, pp. 222–247, 2002.

[34] V. A. Riley, "Human use of automation," Ph.D. thesis, Dept. Psychol. Univ., Minneapolis, MN, 1994.

[35] R. Parasuraman, R. Molloy, and I. Singh, "Performance consequences of automation-induced, complacency," *Int. J. Aviat. Psychol.*, vol. 3, no. 1, pp. 1–23, 1993.

[36] C. Gonzalez and J. Quesada, "Learning in dynamic decision making: The recognition process," *Comput. Math. Organ. Theory*, vol. 9, no. 4, pp. 287–304, 2003.

[37] C. Gonzalez, J. F. Lerch, and C. Lebiere, "Instance-based learning in dynamic decision making," *Cogn. Sci.*, vol. 27, no. 4, pp. 591–635, 2003.

[38] S. Roberts and H. Pashler, "How persuasive is a good fit? A comment on theory testing," *Psychol. Rev.*, vol. 107, no. 2, pp. 358–367, 2000.

[39] J. Gao and J. D. Lee, "Information sharing, trust, and reliance—A dynamic model of multi-operator multi-automation interaction," in *Human Performance, Situation Awareness and Automation: Current Research and Trends*, vol. 2, D. A. Vincenzi, M. Mouloua, and P. A. Hancock, Eds. Mahwah, NJ: Lawrence Erlbaum Associates, 2004, pp. 34–39.

[40] E. Hutchins, "How a cockpit remembers its speeds," *Cogn. Sci.*, vol. 19, no. 3, pp. 265–288, 1995.

Ji Gao received the B.S. degree in mechanical engineering from Tsinghua University, Beijing, China, in 1999, the M.S. degree in high performance computation in engineering systems from the National University of Singapore, Singapore, in 2000, another M.S. degree in mechanical engineering from the University of Iowa, Iowa City, IA, in 2002, and is currently working toward the Ph.D. degree in industrial engineering at the University of Iowa.

She is currently a Research Assistant at the Cognitive Systems Laboratory, University of Iowa. Her research interests are in the area of human–automation interaction and its effect on the emergent properties of group behavior in complex socio-technical systems. Specific interests include modeling dynamic decision making, human-centered system design, and human–automation interaction in decentralized systems.

John D. Lee received the B.A. degree in psychology and the B.S. degree in mechanical engineering from Lehigh University, Bethlehem, PA, in 1987 and 1988, respectively, and the M.S. degree in industrial engineering and the Ph.D. degree in mechanical engineering from the University of Illinois at Urbana-Champaign, in 1989 and 1992, respectively.

He is currently an Associate Professor at the Department of Mechanical and Industrial Engineering and the Director of the Cognitive Systems Laboratory at the University of Iowa, Iowa City, IA. His research enhances the safety and acceptance of complex human–machine systems by considering how technology mediates attention. His research investigates the role of trust in affecting how people manage imperfect automation. Specific research interests include technology-mediated attention, trust in technology, supervisory control, and collision warning systems and driver distraction.