ERGONOMICS, 1996, VOL. 39, NO. 3, 429–460

Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation

BONNIE M. MUIR† and NEVILLE MORAY‡

†36 McMillan Avenue, West Hill, Ontario M1E 4B5, Canada
‡Université de Valenciennes, B.P. 311, F-59304 Valenciennes Cédex, France

Keywords: Trust; Automation; Allocation of functions; Human-computer interaction; Supervisory control

Two experiments are reported which examined operators' trust in and use of the automation in a simulated supervisory process control task. Tests of the integrated model of human trust in machines proposed by Muir (1994) showed that models of interpersonal trust capture some important aspects of the nature and dynamics of human-machine trust. Results showed that operators' subjective ratings of trust in the automation were based mainly upon their perception of its competence. Trust was significantly reduced by any sign of incompetence in the automation, even one which had no effect on overall system performance. Operators' trust changed very little with experience, with a few notable exceptions. Distrust in one function of an automatic component spread to reduce trust in another function of the same component, but did not generalize to another independent automatic component in the same system, or to other systems. There was a high positive correlation between operators' trust in and use of the automation; operators used automation they trusted and rejected automation they distrusted, preferring to do the control task manually. There was an inverse relationship between trust and monitoring of the automation. These results suggest that operators' subjective ratings of trust and the properties of the automation which determine their trust, can be used to predict and optimize the dynamic allocation of functions in automated systems.

1. Introduction

The advent of complex, dynamic, 'intelligent' automation has produced a need to understand the relationship between this automation and the human operators who interact with it. It has been hypothesized that a key factor in this relationship may be the operators' trust in the automation (Muir 1987, Muir 1989, Muir 1994, Sheridan and Hennessy 1984), but until recently, there has been no research to test this hypothesis, nor a theoretical foundation upon which to base empirical studies. Part I of this article (Muir 1994) outlined a theoretical model of human trust in machines; Part II focuses upon the experimental testing of this model and a number of hypotheses proposed there.

The goal of research in the area of human trust in automated systems is to be able to predict and optimize operators' intervention behaviour, on the basis of either (1) their trust, or (2) the properties of the automation which determine their trust. In automated supervisory control environments, operators are given the task of deciding when to intervene and take manual control to optimize system performance. Since intervention in an automated process can have serious consequences, it is important to understand and optimize the operator's decision process. Accordingly, the two experiments reported here were designed to investigate how the properties of automated systems affect operators' trust in the automation and their decision to intervene in an automated process. The experiments address the following issues:

- Feasibility-can operators meaningfully rate their trust in a machine?
- How is trust affected by the properties of the automation?
- How does trust in the automation relate to intervention behaviour?
- How does trust in the automation relate to monitoring behaviour?
- How does distrust of the automation spread in systems?
- How does trust change with experience?
- Are models of interpersonal trust useful bases for models of human trust in machines?

These experiments were completed in 1988, and laid the ground work for extensions by Lee and Moray (e.g. Lee 1992, Lee and Moray 1992).

2. The experimental task

Both experiments reported here were run using a computer-controlled simulation of a milk pasteurization plant, which was created for this purpose. The pasteurization process involves heating raw milk to specified time and temperature criteria to kill bacteria in the milk. The simulation is a medium-fidelity simulation of the high speed, short time, continuous processing method of pasteurizing milk. The operator's task is to control the pasteurization process, maximizing system output within safety constraints.

The mimic display of the simulation is shown in figure 1. Raw cold milk enters the system at variable temperatures, and flows through a pipe at the top left of the screen and is held in the main vat below. If the feedstock pump in the lower left corner is running, the raw milk flows through a passive heat exchanger and into a heater where it is heated by steam supplied by the boiler system. When the milk leaves the heater, it is directed automatically by the three-way valve, depending on its temperature: if the

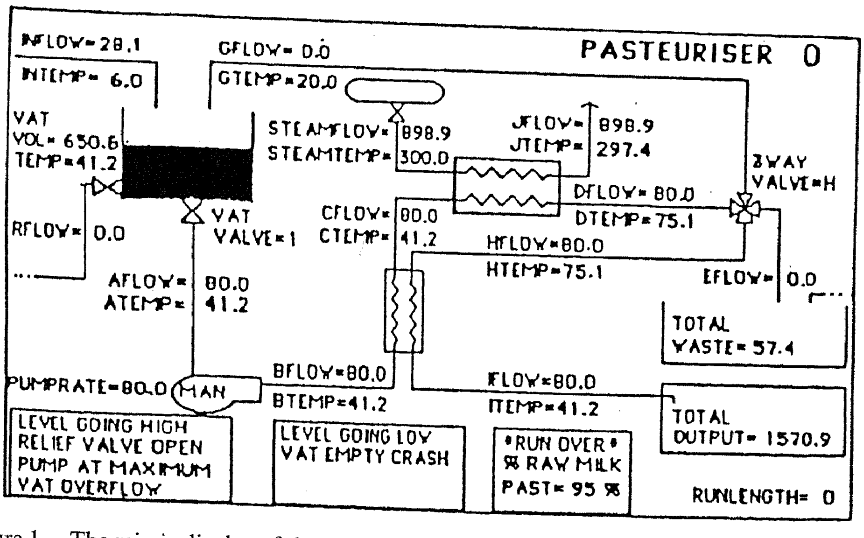

Figure 1. The mimic display of the pasteurizer plant simulation in experiment 1, at the end of a training run. All possible verbal warnings are shown in lower left.

milk is within the acceptable range for pasteurization (70–85° C), it passes through the passive heat exchanger, preheating cold milk, and into the output vat; if the milk is burned (hotter than 85° C), it is directed to the waste vat; if the milk is too cold, it is recirculated back to the main vat to be reheated, thus increasing the vat volume and changing the temperature of the milk in the vat.

From the operators' point of view, the two most important subsystems for controlling the process are the pump subsystem and the heating subsystem. From an experimental point of view, the pump subsystem is the main focus of interest, since this is the automation which is systematically manipulated in the experiments. The pump subsystem is semiautomated; the normal operating mode is automatic, but operators can switch between manual and automatic control whenever they wish by typing the appropriate commands on the control keyboard. In automatic mode, the pump resets its flow rate to match the flow rate of milk into the system. In manual pump mode, operators specify the flow rate through the system by typing in a new pump target, which is echoed on the control screen. In both modes, the pump's response to a new pump target takes the form of a cosine function, changing slowly at first, accelerating, then levelling off. It takes a maximum of about 10 iterations (1 min) to achieve new pump targets, with the current pump rate displayed on the mimic at every iteration, so operators can track the pump's responses over time. The heating subsystem was controlled manually by operators in Experiment 1 and was fully automated in Experiment 2. All other subsystems were accurate and fully automated with no manual override. Realistic heat transfer equations, time lags, and feedback loops make the control task challenging. Operators' scores are displayed at the end of every run.

3. Experiment 1—The nature and dynamics of human trust in machines

3.1. *Introduction*

There are many examples of accidents involving automation, and in some of these we may infer post hoc that operators' errors of mistrust in the automation played a role in the accident, because they trusted the automation either too much or too little (e.g. Wiener and Curry 1980). However, it is one thing to infer trust post hoc, and quite another to ask operators during normal operations to express their trust in the automation they are using. When the present experiment was conducted in 1988, to the best of the author's knowledge, there were no experiments which had explicitly asked operators to rate their trust in automated equipment. Therefore, a fundamental goal of this experiment was to assess the feasibility of studying the construct of trust between humans and machines. Studies of trust between humans (e.g. Rempel *et al.* 1985) indicate that people are able to express quantitatively their trust in others with whom they are involved in stable relationships. The question posed in this experiment was whether operators are able to make subjective ratings of their trust in automation, in the same way, under normal operating conditions.

Automatic controllers vary in quality, some designs being better than others at performing a given task, so it is inappropriate for operators to trust them all equally. Even automatic controllers of the same design may have idiosyncratic properties or 'signatures' (Woods 1986); they are not faulty in the sense of failing or breaking down; rather, they behave in certain systematic, less-than-ideal ways. For example, one automatic pump in a system may adjust more sluggishly than another to changes in system states. In this experiment, nine different pumps, varying systematically in the properties of their control and display responses and ranging from excellent to disastrous system controllers, were simulated to determine how the properties of the automation affected operators' trust in the automation.

When operators rate their trust, upon what aspects of the automation are they basing these ratings? In Part I of this paper (Muir 1994), a model of the meanings of trust in machines was presented, which was based upon Barber's (1983) taxonomy of the three meanings of trust in interpersonal relationships. According to this model, trust in automation is a composite expectation of (1) the operator's general expectation of the *persistence* of the natural physical order, the natural biological order and the moral social order, (2) a specific expectation of the *technical competence* of the automation and (3) a specific expectation of the *fiduciary responsibility* of the automation. This experiment focused on the latter two meanings, which refer to properties of the automation as bases for overall trust. When it was undertaken, this was exploratory research in a new field; with the sensitivity of the response (dependent) variable of trust unknown, it was prudent to have a simple, external, clearly manipulated explanatory (independent) variable, such as system properties. Accordingly, the properties of the automation were chosen to differ in a simple, transparent way from an explicit standard. (It was not the goal of these experiments to determine whether operators can detect certain system properties. Rather, it was to determine how their trust in and use of the automation is affected, given that operators perceive certain properties). Barber's aspect of persistence emphasizes the importance of the evaluator's stable disposition and when some general groundwork has been done, future research can investigate such individual differences in human-machine trust. In this experiment, operators completed subjective ratings of different automatic controllers' competence and responsibility. These were evaluated as predictors of operators' ratings of overall trust in the automatic controllers. Given the relatively simple automatic controllers in the simulation, it was predicted that competence would be the more applicable of the two meanings of trust tested. However, if the *absolute magnitude* of variance in overall trust accounted for by competence and responsibility is low, then operators have in mind some other meaning of trust not included in the proposed model and this meaning remains to be specified.

In Muir (1994), a model of how operators' trust might change with experience was proposed. This model, based upon a model of the dynamics of interpersonal trust (Rempel *et al.* 1985), was also tested in this experiment. Operators made subjective ratings of the predictability and dependability of the automatic controllers, and their faith in them. The relative importance of each of these factors was evaluated at different points in the operators' experience on the system. Predictability should account for more of the variance in novices' overall trust than dependability or faith, to the extent that the human-human model of trust development transfers to the human-machine relationship. However, as they gain more experience, the importance of dependability as a predictor should increase. Finally, faith in the automation should emerge as the best predictor of expert operators' overall trust in the automation.

The question of how distrust may spread within automated systems is of considerable theoretical interest and practical importance. If operators come to distrust one automatic component of a system, will their distrust generalize to diminish their trust in other components? If so, which ones? Distrust might spread between components which are structurally related, functionally related, or causally related (Sheridan and Hennessy 1984), or spread indiscriminately over the whole system, flooding it so that trust is reduced in all parts of the system. Two possible paths of the spread of distrust were tested in this experiment: (1) between two different functions of a single subsystem (structurally related) and (2) 'flooding' between two subsystems which are structurally and functionally independent. The first path was tested by having operators make separate ratings of their trust in the control function and the display function of each of the manipulated pumps and then examining whether distrust in the pumps' display properties affected their trust in the pumps' control properties and vice versa. To test the flooding hypothesis, operators rated their overall trust in the pumps as well as their overall trust in another independent and consistently accurate automatic controller in each system (the three-way valve), to determine whether distrust in the poor automatic controllers would generalize to reduce trust in the independent, accurate automatic controllers.

3.2. *Method*

3.2.1. *Participants*: Six male graduate students at the University of Toronto volunteered to participate. They were paid an hourly rate plus a bonus contingent on performance.

3.2.2. *Experimental task*: Operators controlled the pasteurizer plant simulation described in Section 2. Operators could intervene to take manual control of the automatic pump as often and for as long as they wished, in an effort to maximize system performance. The operator's score was calculated as the percentage of raw milk entering the system that was pasteurized, less a penalty of 30% if the vat volume was outside the safe range at the end of a run.

3.2.3. *Experimental conditions*: Only the properties of the pump were manipulated in this study; all other subsystems were consistently exact (accurate). Both the control and the display properties of the pump were manipulated to be either exact, have a constant proportional error, or have a variable error (described in more detail below). The nine task conditions thus formed are summarized in table 1, along with an example. It is helpful to think of these conditions as being formed in three sequential steps: (1) what the pump is *requested* to do (i.e. the pump target), (2) what the pump *actually* does (the control manipulation) and (3) what the pump *claims* to be doing (the display manipulation).

For the control manipulation, the pump's asymptotic pump rate in response to a new pump target was either exactly the same as the pump target requested, or manipulated to be 10% higher than the pump target requested, or manipulated to be a value selected randomly from a Gaussian distribution centred on the requested pump target, with a standard deviation of 2% of the pump target value. These manipulations were the same whether the pump was in manual or automatic mode.

With respect to the display manipulation, the pump displayed its current pump rate, updated at every iteration, on the mimic diagram (figure 1). The displayed value represents the flow rate that the pump claims to be producing. The properties of the pump rate display were manipulated in the same way as the pump's control properties, except that the display manipulation was performed upon the pump's *manipulated* output, not upon the requested pump target. Thus, the pump rate display on the mimic could show either an honest, faithful display of the manipulated pump rate, or a rate 10% higher than the manipulated pump rate, or a rate with a variable Gaussian error (mean = 0, standard deviation = 2%) added at every iteration to the

Table 1. Summary of the nine experimental conditions in experiment 1. PT = pump target, RNV = random normal variable, C = experimental condition. Sample values of the manipulated pump target and pump rate display given in response to a request for a pump target of 20 litres/minute are shown in square brackets.

	Display manipulation (of displayed pump rate)		
Control manipulation (of commanded pump target) (e.g. 20)	honest (manipulated PT + 0)	constant error added (manipulated PT + 10%)	variable error added (manipulated PT + RNV)
exact (commanded PT + 0) [20]	C1 [20]	C2 [22]	C3 [21.3]
constant error added (commanded PT + 10%) [22]	C4 [22]	C5 [24.2]	C6 [21.8]
variable error added (commanded PT + RNV) [20.4]	C7 [20.4]	C8 [22.4]	C9 [20.7]

manipulated pump rate. The word 'honest' is used to describe the first type of display, rather than 'accurate' or 'exact', because the latter two terms could be misleading in Conditions 4 and 7 in which the display accurately reflects the actual (manipulated) output, but is not accurate in relation to the requested pump target.

The three control and the three display manipulations were completely crossed, yielding nine experimental conditions. Each condition was randomly mapped onto a separate pasteurizer system in the plant, so there were nine pasteurizers in total in the plant, with the pump in each system having a unique combination of control and display properties. There was sufficient information on the mimic and control screens to allow operators to evaluate the accuracy of each pump's control and display properties in both manual and automatic modes.

3.2.4. *Subjective rating scales*: After every run, operators completed a set of subjective rating scales for the pump subsystem (in general, including its automatic and manual modes). 100-mm scales were used, with the poles labelled 'none at all' or 'not at all' on the left to 'extremely high' on the right. To minimize the variance in operators' interpretations of the qualities being rated, each scale was accompanied by a brief written definition of the quality. Operators completed the following scales for the pump:

- *competence* (i.e. to what extent does the pump perform its function properly?). [A further verbal clarification was given: 'To what extent does it produce the requested flow rates?']
- *predictability* (i.e. to what extent can the pump's behaviour be predicted from moment to moment?).
- *dependability* (i.e. to what extent can you count on the pump to do its job?).
- *responsibility* (i.e. to what extent does the pump perform the task it was designed to do in the system?) [A further verbal clarification was given. 'To what extent does it maintain system volume?']
- *reliability over time* (i.e. to what extent does the pump respond similarly to similar circumstances at different points in time?).

- your degree of *faith* that the pump will be able to cope with other system states in the future.
- your degree of *trust* in the pump to *respond* accurately.
- your degree of *trust* in the pump's *display*.
- your *overall degree of trust* in the pump.

Operators also completed an identical set of scales for the consistently accurate three-way valve.

3.2.5. *Training procedure*: Operators were trained to control the simulated process on a training pasteurizer whose control and display properties were exact. The rigorous training regimen was designed to allow operators to become highly skilled in operating the system in manual pump mode, in automatic pump mode, and in switching pump modes, to more closely resemble the experience of real-world operators. Operators were trained individually by the experimenter, in one-hour sessions, consisting of 4 runs of 80 iterations each, about 10 minutes per run in real time. Five stages of training emphasized, respectively: introducing the operator to the system using a detailed training manual, practising manual pump control, practising switching between manual and automatic modes, practising automatic control and free choice of automatic or manual control. Subjective rating scales were completed after every run. Training was completed when the operator achieved a mean score of at least 80% for a session in Stage 5. The minimum required training time was 7 1-hour sessions. Operators actually took from 8–16 sessions. The mean number of sessions required was 11·5, representing about 7·2 hours real time on the control task.

3.2.6. *Experimental procedure*: After training, operators were told that they would be operating the nine pasteurizers in the plant, but no information was given about their properties. Operators worked for one 1-hour session on each pasteurizer, controlling the simulation for a run of 80 iterations and then completing the subjective ratings of the pump and three-way valve, four times per session.

3.3. *Results*

3.3.1. *Subjective ratings of trust*: The ratings scales were scored by measuring the number of mm from zero at the left pole, yielding scores from 0–100. Separate analyses of variance were performed on the three subjective ratings of trust: overall trust, trust in the control properties and trust in the display properties of the pumps and three-way valves. There were four factors in each analysis: control properties (exact, constant error, variable error); display properties (honest, constant error, variable error); runs within sessions (Runs 1, 2, 3, 4); the component being rated (pump, three-way valve). Complete results of these analyses are reported in Muir (1989).

3.3.1.1. *Overall trust*: Figure 2 shows operators' overall trust in the pumps and three-way valves in the nine plant pasteurizers as a function of the control and display properties of the pumps. Turning first to the effect of the control manipulation, ratings of overall trust were affected by the pumps' control properties, but this effect interacted with the component being rated and the runs factor, $F(6, 30) = 2{\cdot}43$, $p < 0{\cdot}05$. With respect to the pumps, post hoc multiple comparisons (Newman-Keuls tests at $a = 0{\cdot}05$) revealed that operators trusted the exact control pumps more than

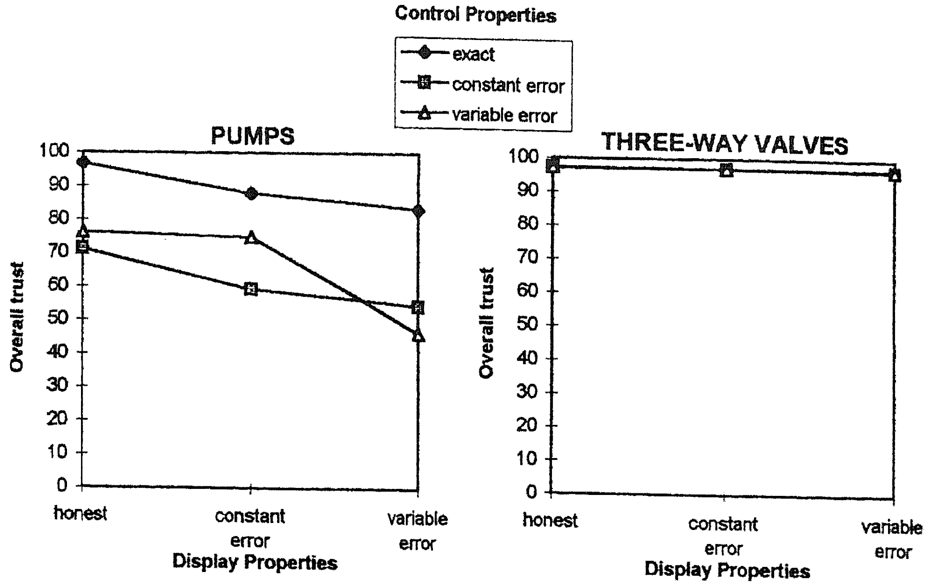

Figure 2. Mean ratings of overall trust in the pumps and three-way valves, pooled over operators and runs, as a function of the control and display properties of the pumps.

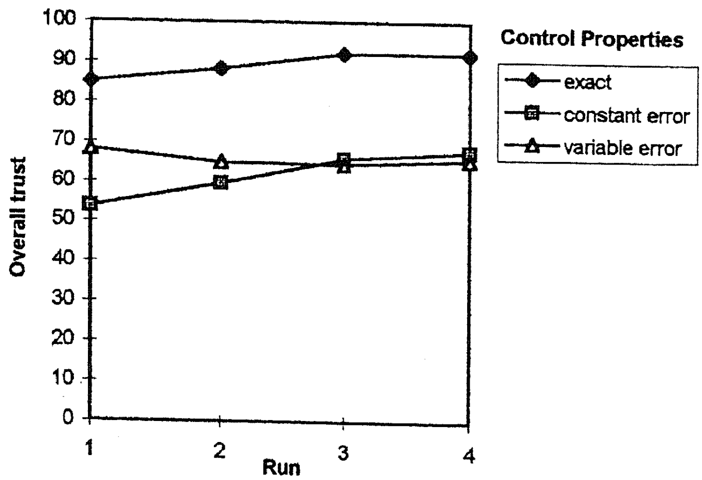

Figure 3. Mean ratings of overall trust in the pumps, pooled over display properties and operators, as a function of the control properties of the pumps and operators' experience.

the pumps with a constant control error or a variable control error and their trust in the latter two did not differ. Simple effects analyses showed that operators' trust in the constant control error pumps grew with experience, $F(3, 15) = 9{\cdot}32$, $p < 0{\cdot}005$, but was more stable for the exact control and variable control error pumps, $F(3, 15) = 2{\cdot}69$, $0{\cdot}05 < p < 0{\cdot}10$ and $F(3, 15) < 1$, respectively (figure 3). In contrast, overall trust in the three-way valves was consistently high (figure 2) and post hoc

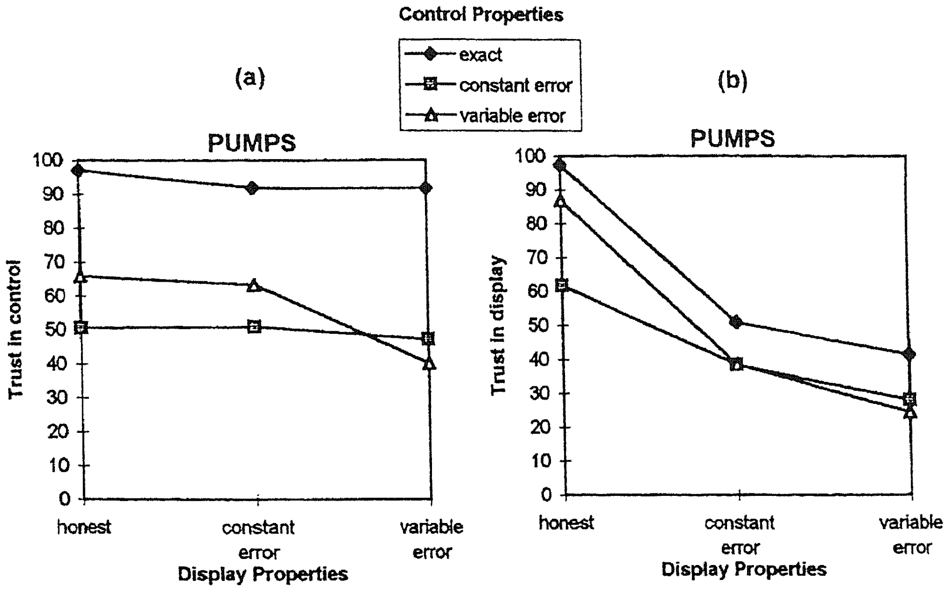

Figure 4. (a) Mean ratings of trust in the control properties of the pumps. (b) Mean ratings of trust in the display properties of the pumps. Both are pooled over operators and runs, as a function of the control and display properties of the pumps.

multiple comparisons confirmed that it did not vary as a function of the pumps' control properties or practice.

The display manipulation also affected overall trust in the pumps, but did not affect trust in the three-way valves, $F(2, 10) = 7{\cdot}37$, $p < 0{\cdot}01$. Post hoc comparisons revealed that operators trusted the pumps with honest displays more than those with constant display errors, which in turn were trusted more than those with variable display errors (means = 81·4, 74·1, 61·5, respectively). In contrast, overall trust in the three-way values was consistently high and unaffected by the pumps' display properties.

The control and display manipulations did not interact, $F(4, 20) < 1$, suggesting that the control and display properties had additive effects on operators' ratings of overall trust.

3.3.1.2. *Trust in control*: Figure 4(a) shows operators' trust in the control properties of the pumps in the nine plant pasteurizers. A complete analysis of variance was performed on these data, but only the results relevant to the generalization of distrust are reported here. Simple effects analyses showed that operators' trust in the pumps' control properties was affected by the display manipulation, $F(2, 10) = 7{\cdot}99$, $p < 0{\cdot}01$. Thus, even though the control and display manipulations were independent, the *display* properties of the pumps affected operators' trust the pumps' *control* properties.

3.3.1.3. *Trust in the display*: Figure 4(b) shows operators' trust in the displays of the pumps in the nine plant pasteurizers. Trust in the pumps' displays was affected by both the control and the display manipulation, but the effects were not additive. The effect of the display manipulation was fairly uniform for the three control conditions, except that the honest display/constant control error pump was rated lower than the other two honest display pumps. Simple effects analyses confirmed that trust in the display was reliably affected by the display manipulation for the exact control pumps, $F(2, 10) = 9{\cdot}26$, $p < 0{\cdot}01$ and variable control error pumps, $F(2, 10) = 11{\cdot}02$, $p < 0{\cdot}005$, but to a lesser extent for the constant control error pumps, $F(2, 10) = 3{\cdot}06$, $0{\cdot}05 < p < 0{\cdot}10$. It is not surprising that the control manipulation influenced operators' ratings of trust in the pumps' displays: the display manipulation was performed upon the pumps' *manipulated output*, and so display inaccuracies are compounded with control inaccuracies in the two inexact control conditions, but comprise a single inaccuracy in the exact control condition.

3.3.2. *The meaning of trust in machines*: To test the proposed model of the meanings of trust in a machine (Muir 1994), based on Barber's (1983) model, simple linear regression analyses of competence and responsibility on overall trust were performed. The results appear in table 2. Pooled across all experimental conditions, competence was the better single predictor of trust, accounting for 81% of the variance in overall trust, compared to 76% for responsibility. (Mallows' $C(P)$ statistic was also used to evaluate predictors in all of the regressions reported, but in no case did it conflict with the R^2 criterion, so it is not reported here.) Although responsibility accounted for almost as much variance as competence, competence was consistently dominant in the stepwise regression analyses. A stepwise procedure on the pooled data selected a single factor model as the best model, with competence as the predictor and the resulting regression model was highly reliable, $F(1, 214) = 915{\cdot}8$, $p < 0{\cdot}0001$. The results for the individual experimental conditions followed the same pattern as the pooled analysis. In eight conditions, competence accounted for more variance than responsibility and adding responsibility to the model made little improvement in the R^2 over competence alone. In seven conditions, the stepwise procedure selected a single factor model, with competence as the predictor and the regression equations were similar and highly reliable.

Table 2. Summary of analyses of the meanings of trust. C = competence, Rs = responsibility, T = overall trust. The best single predictor of overall trust in each condition is underlined.

Experimental Condition	R^2 Competence	R^2 Responsibility	R^2 Competence, Responsibility	Stepwise Best Predictor	Stepwise Best Model	F(Best Model)	p
Pooled	0·81	0·76	0·81	C	$T = 14{\cdot}1 + 0{\cdot}8C$	915·8	< 0·0001
1	0·76	0·65	0·76	C	$T = 17.1 + 0{\cdot}8C$	70·28	< 0·0001
2	0·75	0·74	0·77	C	$T = 5{\cdot}5 + 0{\cdot}9C$	66·80	< 0·0001
3	0·79	0·72	0·79	C	$T = 12{\cdot}8 + 0{\cdot}8C$	81·04	< 0·0001
4	0·20	0·12	0·21	C	$T = 49{\cdot}1 + 0{\cdot}3C$	5·57	< 0.05
5	0·92	0·89	0·92	C	$T = 5{\cdot}6 + 0{\cdot}9C$	269·19	< 0·0001
6	0·87	0·92	0·92	Rs	$T = -3{\cdot}2 + 1{\cdot}0Rs$	243·57	< 0·0001
7	0·91	0·81	0·95	C, Rs	$T = 22{\cdot}7 + 1{\cdot}5C - 0{\cdot}7Rs$	179·04	< 0·0001
8	0·68	0·67	0·69	C	$T = 41{\cdot}3 + 0{\cdot}5C$	46·66	< 0·0001
9	0·92	0·79	0·92	C	$T = 7{\cdot}84 + 0{\cdot}8C$	267.43	< 0·0001

In a *relative* sense, competence was a better predictor of overall trust than responsibility. Moreover, competence as also a good predictor in an *absolute* sense, accounting for 81% of the variance in trust pooled over conditions and 20–92% in individual conditions. The R^2 of 0·20 for Condition 4 is far below that of the other conditions, due to two anomalous responses from a single operator. If the data for Condition 4 are excluded, competence accounted for 68–92% of the variance in trust in the other eight conditions.

3.3.3. *The development of trust in machines*: According to Muir's (1994) model, based on Rempel *et al.*'s (1985) stage model, predictability should be the best predictor of overall trust early in an operator's experience, followed later by dependability and then faith. To test whether this order of emergence applied to operators' trust in the pumps in this task, regressions of the form:

$$\text{Trust} = \text{Predictability} + \text{Dependability} + \text{Faith}$$

were performed on the data at three points during the operators' experience: the first training session, the last training session and the last run of the experimental session in Condition 1 (a span of about 5 weeks in real time). The pumps in these three sessions were identical, with exact control and display properties. The results of these analyses are shown in table 3. The R^2 and Mallows' $C(P)$ criteria were in agreement in most cases so only R^2 is reported here. The regression equations are not included in the table because they are probably highly specific to referents and tasks and in this analysis it is the relative importance of the three factors that is of primary interest. Considering first the R^2 values, in the first training session, faith accounted for the most variance in trust in the pump, followed by dependability, then predictability. This ordering is opposite to that predicted by Rempel *et al.*'s model, but this conclusion must be mitigated by the fact that the stepwise regression model falls slightly short of statistical significance. For the last training session, dependability accounts for the most variance; however, the R^2s for dependability and predictability are quite close, and both are greater than that for faith, a pattern which is consistent with Rempel *et al.*'s second stage of trust. For the plant session, predictability accounts for the most variance, but once again, the R^2s for predictability and dependability are close and both are greater than that for faith, consistent with Rempel *et al.*'s second stage of trust. The results of stepwise procedure show the three factors emerging as predictors in the opposite order to that predicted, however, these data possess a high degree of multicollinearity, so these regression models must be

Table 3. Summary of analyses of the development of trust. P = predictability, D = dependability, F = faith. The best single predictor of overall trust in each session in underlined.

Session	R^2 Predictability	R^2 Dependability	R^2 Faith	Stepwise Best Predictors	F(Best Model)	p
First training	0·39	0·58	<u>0·63</u>	F	6·75	$< 0{\cdot}06$
Last training	0·96	<u>0·99</u>	0·46	D, F	1234·78	$< 0{\cdot}0001$
Plant (Experimental)	<u>0·78</u>	0·72	0·46	P	14·57	$< 0{\cdot}02$

interpreted cautiously. (For example, in the plant session, the intercorrelations were $r(4) = 0{\cdot}95$, $p < 0{\cdot}01$ for predictability and dependability, $r(4) = 0{\cdot}73$, n.s. for predictability and faith and $r(4) = 0{\cdot}90$, $p < 0{\cdot}05$ for dependability and faith.) The relative values of the R^2s are more useful, and they show a pattern that is fairly consistent with Rempel *et al.*'s model, with the exception of the faith factor: there was a tendency for faith to emerge before predictability and dependability, but it had not emerged as an important predictor by the third session examined.

To test whether the other ratings gathered in this study might be better predictors of developing trust than Rempel *et al.*'s predictors, all the regressions were repeated using the full model:

$$\text{Trust} = \text{Predictability} + \text{Dependability} + \text{Faith}$$
$$+ \text{Competence} + \text{Responsibility} + \text{Reliability}.$$

In no case did the added factors outperform Rempel *et al.*'s factors. The fact that the R^2s for their three predictors were quite high and no other factors were consistently better, provides some support for the stages of their model.

All of the above regression analyses were repeated using a shorter time horizon (within experimental sessions), but the results were noisy, suggesting that longer time horizons are more suitable for testing models of trust development.

3.3.4. *Performance measures*: Separate analyses of variance were performed on operators' scores, the number of manual pump control actions and the proportion of time spent in automatic pump mode. There were three factors in each analysis: the control properties of the pump (exact, constant error, variable error); the display properties of the pump (honest, constant error, variable error); runs within sessions (Runs 1, 2, 3, 4). Reported means are pooled over operators and runs, in Conditions 1–9 respectively.

3.3.4.1. *Scores*: Mean scores were high in all experimental conditions (98·9, 98·7, 97·0, 98·4, 95·8, 97·3, 96·3, 95·6, 97·7), and were not affected by the control properties or the display properties of the pumps, $F(2, 10) = 1{\cdot}53$, n.s., $F(2, 10) < 1$, respectively. The only significant effect was an interaction between the control manipulation and runs, $F(6, 30) = 3{\cdot}06$, $p < 0{\cdot}02$; scores improved with practice on the pumps with a constant control error, $F(3, 15) = 3{\cdot}37$, $p < 0{\cdot}05$, but remained more stable for the accurate pumps, $F(3, 15) = 2{\cdot}08$, n.s., and the variable control error pumps, $F(3, 15) = 2{\cdot}80$, $0{\cdot}05 < p < 0{\cdot}10$.

3.3.4.2. *Pump control actions*: The mean number of times operators adjusted the pump target (28·4, 28·9, 29·8, 30·0, 31·1, 34·8, 28·9, 29·3, 30·0) did not differ as a function of the pumps' control or display properties, $F(2, 10) = 2{\cdot}76$, n.s., $F(2, 10) = 1{\cdot}38$, n.s., respectively, but tended to decrease with practice in all experimental conditions, $F(3, 15) = 3{\cdot}14$, $p < 0{\cdot}06$.

3.3.4.3. *Proportion of time in automatic pump mode*: The mean proportion of time spent in automatic pump mode was low in all experimental conditions (0·12, 0·10, 0·16, 0·12, 0·14, 0·12, 0·09, 0·11, 0·09) and did not differ as a function of the control properties or the display properties of the pump, both F's$(2, 10) < 1$, or practice, $F(3, 15) = 1{\cdot}38$, n.s.

3.3.5. *Correlation between performance measures and subjective ratings of trust*: The correlation between overall trust in the pumps and performance scores was statistically significant, $r(214) = 0{\cdot}22$, $p < 0{\cdot}001$, but scores accounted for only 5% of the variance in trust. Thus, scores may play a small role in determining operators' trust, but other factors, such as the properties of the automation, must contribute as well. Operators' trust did not correlate with the number of adjustments they made to the pumps, $r(214) = -0{\cdot}10$, n.s. The correlation between trust and use of the automatic pumps was statistically significant, $r(214) = -0{\cdot}16$, $p < 0{\cdot}02$, however, trust accounted for only 2% of the variance in use of the automatic pumps, and is in the opposite direction to that predicted. These correlations must be interpreted with caution because the three performance measures were truncated in range.

3.4. *Discussion*

It is clear from the results of Experiment 1 that the construct of human trust in machines is amenable to study in an experimental situation. Operators were able to make subjective ratings of their trust in the pumps, both the exact and the inexact ones. Their ratings were sensitive to the properties of the pumps, and related in a sensible way to these properties. Operators did sometimes express difficulty in making their ratings, but, unexpectedly, it was not rating the accurate pumps that was problematic: rather, operators struggled most in deciding on the relative importance of the constant and variable errors. These verbal expressions, together with their sensible ratings, suggest that operators made a conscientious, careful effort to perform a difficult and onerous task. Operators' overall trust in the pumps was affected by both the control and the display properties of the pumps. Their trust ratings indicate that they were able to detect and discriminate between the two manipulated properties in the same subsystem and took these properties into account in an additive way in their expressions of trust. This is consistent with other research that shows that people, even domain experts like medical clinicians, use simple additive models, even when more complex ones are indicated (Goldberg 1968).

An interesting result of this study was that the two types of control error had similar effects on operators' trust. Averaged over their experience, a small variable error had approximately the same psychological impact as a relatively large constant error. However, a closer examination of the development of trust in the three control conditions revealed an important difference. Initially, trust in the constant error pumps was much lower than trust in the variable error pumps, probably because operators were comparing the absolute magnitude of control errors and this was much larger in the constant error pumps. However, with experience, operators' trust in the constant error pumps grew to the same level as that for the variable pumps, probably because they perceived the constancy of the error and learned that they could easily compensate for it by requesting pump targets 10% lower than they actually wanted. In contrast, there was no strategy for compensating for variable error and operators' trust in the variable error pumps remained stable and low. This pattern of results suggests that the control error's consequences for system control are an important determinant of trust: if operators can discover a systematic bias in a component and if they can find a control strategy to compensate for it, their trust in that component may grow. On the other hand, if operators cannot compensate for the error, as with the variable error pumps, then their trust will remain low. Operators seem to be willing under some circumstances to adjust their trust in light of further evidence. This finding supports the hypothesis in Part I of this paper (Muir 1994) that operators must be given plenty of experience with the automation to discover its properties, calibrate their trust and find ways to exploit it.

With respect to the generalization of distrust in systems, the results of this study suggest that distrust can spread between two separate functions within a common physical component. The strongest evidence for this is that the display properties of the pumps affected operators' trust in the pumps' control properties. It is important to note that operators could perform several independent checks on the accuracy of the pumps' control and display properties using the information available, so the observed generalization of distrust is unlikely to be attributable to confusion. This result has important implications for real-life automated systems in which individual structures or subsystems may perform many functions, which may be quite independent. If operators' distrust of a particular function spreads to other functions performed by the same subsystem, this may lead to unwarranted distrust, unnecessary monitoring and inappropriate manual override of perfectly good functions in a subsystem. Although contagious, distrust in the automatic pumps did not spread indiscriminately, flooding the whole system: operators' trust in the automatic three-way valves was consistently high and unaffected by the properties of the pumps. Thus, in this experiment, trust was specific to physical subsystems.

The results of this experiment provided some empirical support for the integrated model of trust presented in Part I of this paper (Muir 1994). The tests of the meaning of trust in a machine showed that the expectation of competence best captures what operators mean when they say they trust a machine. This result is consistent with our intuition; after all, we create machines for the specific purpose of performing a function for us. Competence was operationally defined here as the extent to which the pump performs its function properly, which is to produce the requested flow rates. In rating a pump's competence, operators are considering this: given its design, does it produce the flow rates I ask for? Responsibility was operationally defined as design-based responsibility: the extent to which the pump performs the task it was designed to do in the system, which is to maintain system volume. This is a more general property than competence. A pump is designed to control flow rates effectively and efficiently, but many different designs are possible. For instance, the pumps used in this study could have been designed to change pump rates faster or more 'intelligently' than they did. Thus, the responsibility scale invokes a comparison of the pump's particular design with other possible designs. That overall trust was related more to competence than to responsibility implies that operators used a concrete criterion of performance, conditionalizing their trust on the pump's actual behaviour in relation to the way it was designed to work, rather than on an abstract comparison of its actual behaviour with ideal behaviour. The meanings of competence and responsibility, as they were defined for this study, were similar and they were rated fairly similarly by operators. This is probably due to the relatively simple, transparent automation studied. The relative importance of competence and responsibility may be task specific, with responsibility dominating in more complex, 'intelligent', or less transparent automation, whose competence is more difficult to assess. However, the most important result of this analysis is that, in an absolute sense, both competence and responsibility accounted for a remarkably high proportion of the variance in overall trust, indicating that both are important components of human trust in machines.

The results of this study also provided some empirical support for the second dimension of Muir's (1994) integrated model of trust, namely the development of trust in machines as a result of experience. Predictability, dependability and faith accounted for high proportions of variance in trust over time, indicating that they capture important aspects of the development of trust, but the data on order of emergence are noisier than expected. In particular, there was a suggestion in the data that faith may be a determinant of trust early in operators' experience. It is likely that this result reflects a different meaning of faith than that which Rempel *et al.* hypothesized to characterize mature interpersonal relationships based on intrinsic motivation. Here faith was operationally defined as a belief that the pump would be able to cope with other system states in the future. Given this definition, the faith that novice operators expressed may have been based upon the general expectations about simulations and experimental tasks that they brought to the situation. If so, this would probably be reflected in early ratings. The other deviation with respect to faith is that it had not emerged as an important predictor of overall trust by the end of the experiment. One reason for this may be that Rempel *et al.*'s model was designed to describe close dyadic relationships, evolving slowly over time and experience. Their model may transfer better to human-machine relationships of the same nature, such as those between expert operators and real-world automated systems. Another reason is likely to be the high degree of multicollinearity in the subjective ratings, an outcome also observed in the data of Rempel *et al.* and Lee and Moray (1992). The high degree of variability accounted for by Rempel *et al.*'s three factors warrants retaining the model as a theoretical foundation for further research, but consideration will need to be given to the methodological issues mentioned above.

The relationship between operators' trust in and use of the automatic pumps could not be assessed meaningfully in this study because operators showed a consistent preference for manual control. Operators reported that they felt they were 'chasing' the automatic pump, quickly trying to adjust the heating variables to compensate for automatically controlled changes in pump rate. To maximize their performance under these conditions, operators chose manual pump mode. Even though they could not escape from the pump's error in manual pump mode, they did have a higher degree of control because they could compensate for the pump's errors by adapting their manual control strategies and could plan and execute coordinated pump and heating changes. This finding has an important implication for laboratory research in this area: to elicit supervisory control behaviour, it is necessary to use a fully automated system with automatic controllers which are effective not only at their own local level, but also at the general level of contributing to good system performance, or operators will override the automation. To make any inferences about real supervisory control systems, where automatic control is the preferred and usual mode, it is necessary to use a laboratory task where automatic control is also the preferred and usual mode.

4. Experiment 2—The relationship between the properties of the automation, trust, and human intervention

4.1. *Introduction*

The primary purpose of this experiment was to determine how the properties of the automation affect operators' trust and intervention behaviour in the context of a higher-fidelity supervisory control task which elicited supervisory rather than manual control strategies. Accordingly, the pasteurizer simulation was modified to be fully automated and the automation itself was improved so that the system would run very effectively in automatic mode, without any intervention from the human supervisor.

An intriguing finding of experiment 1 was that a small, variable control error in the pump reduced trust as much as a larger, constant one. However, the two factors—*magnitude* of control error and *variability* of control error—were confounded in that study, so the importance of each could not be ascertained. Other researchers have also identified these two factors as important determinants of trust: Rempel and Holmes (1986) claim that behaviour must be both positive and consistent for interpersonal trust to grow and Sheridan and Hennessy (1984) have suggested that human-machine trust is influenced by both the desirability and predictability of the consequences of machine behaviour. Experiment 2 was designed to examine the role of desirable and consistent machine behaviour in determining operators' trust. Must machine behaviour be perfectly desirable and consistent to foster trust? Is there any tolerance in trust for control error, even an error or disturbance too small to affect overall system performance? These questions were addressed in this experiment by systematically varying the desirability and consistency of the automation in a completely crossed design. Five magnitudes of control error in the automatic pump were chosen to span the range between what would be considered a very good automatic pump and a very poor one. In addition, the automatic pumps' responses could have a small, constant, Gaussian disturbance added, or be consistent, with no disturbance. The degree of disturbance added was so small that it did not, over the course of a fully automated run, degrade the performance of the variable automatic pumps compared to the consistent ones. This permitted a test of the influence of variability per se on operators' trust in and use of the automatic pumps.

The results of experiment 1 indicated that the competence of the automation is an important determinant of operators' trust. Since both the magnitude and variability of the control error in the automatic pumps represent aspects of competence, it was predicted that the manipulation of these two factors in experiment 2 would reliably influence operators' trust in the automatic pumps. While it was predicted that operators would trust the good automatic pumps (with little or no control error) and distrust the poor ones (with large control errors), the exact nature of the relation could not be predicted (for instance whether trust would fall off gradually as the magnitude of control error increased or precipitously at a particular magnitude). Similarly, it was predicted that operators would trust the consistent automatic pumps more than the variable ones. However, this effect was predicted to interact with the magnitude of the control error factor because the standard deviation of the variable disturbance was constant, which made the disturbance relatively larger for the automatic pumps with small control errors. For example, a poor automatic pump that pumps 40 litres/minute (l/min) more than it should, is not much worse for having had ± 1–2 l/min added, but an automatic pump that pumps only 5 l/min more than it should is relatively more influenced by the ± 1–2 l/min. Consequently, it was predicted that variability would reduce trust more in the automatic pumps with a small (or no) control error than in those with a larger control error.

Based on the evidence showing that operators will override automation which they believe to be inadequate for the job (e.g. Wiener and Curry 1980, Woods *et al.* 1987), it was predicted that operators would use the better automatic pumps more than the poorer ones. However, as with trust, no specific prediction could be made about whether the time spent in automatic mode would gradually diminish over the range of control errors studied, or whether use would be nonlinear, with operators completely rejecting the use of the automatic pumps which exceeded some level of control error. The variability manipulation was chosen so that variable automatic pumps would score as well as consistent ones, hence it was predicted that the variability manipulation would not influence operators' allocation strategies.

The predictions for trust and use of the automatic pumps were similar, so a positive correlation was predicted between trust in and use of the automation. This correlation has been found by Zuboff (1988) in her case studies examining the introduction of new automation. The present study provided an experimental test of Sheridan and Hennessy's (1984) hypothesis, expressing current conventional wisdom, that beyond some criterion of distrust, operators will reject the automation and prefer to perform the task manually.

The relationship between operators' trust and their monitoring behaviour was also examined in this experiment. Sheridan and Hennessy (1984) have suggested that operators will spend more time monitoring automatic controllers they distrust. Rempel *et al.*'s (1985) model predicts the other side of the coin, that operators will cease to monitor the behaviour of trusted automation, relying instead on their attribution of dependability or their faith in it as a basis for their trust. Thus, an inverse relationship between monitoring and trust was predicted in this study.

Finally, this experiment permitted further examination of two issues considered in experiment 1. First, it provided another opportunity to examine the development of operators' trust as a result of experience on different systems, permitting a test of the hypothesis that distrust may be more resistant to change than trust (Muir 1994). Secondly, it allowed another examination of the spread of distrust, this time between independent but highly similar systems in a plant.

4.2. *Method*

4.2.1. *Participants*: Six male graduate students at the University of Toronto volunteered to participate. None had participated in experiment 1. They were paid an hourly rate plus a bonus contingent on performance.

4.2.2. *Experimental task*: Operators controlled the pasteurizer simulation described in Section 2, modified as follows. The process was fully automated and ran effectively without any intervention from the operator. All subsystems were completely automated, with no manual override, except for the pump subsystem; the pump's default mode was automatic but operators could override the automatic pump and operate it manually if they judged this to be necessary to optimize system performance. All flow rates except the inflow were removed from the mimic diagram to obtain a monitoring measure; operators could monitor the accuracy of the automatic pump (while in automatic pump mode) by typing a request for the current pump rate to be displayed on the control screen and comparing this to the inflow.

The operators' goal was to maximize their score in the *easiest* possible way, which was in automatic pump mode whenever possible. The maximum total score was 200 points: the productivity score was the percentage of available raw milk that was pasteurized; for the safety score, 1 point was awarded for every iteration in which vat volume was in the safe range, to a maximum of 100 points. The extent to which the operator met the productivity and safety objectives in the easiest possible way (i.e. in automatic pump mode) was scored by incurring a small penalty for use of the manual pump—for every iteration in manual pump mode, the operator lost 1/10 of a point, to a maximum of 10 points per run. This penalty was chosen to be large enough to discourage operators from competing with the automatic pump by maintaining constant manual control, but small enough not to deter manual control if this was truly warranted. The system with an exact automatic pump achieved a mean score of 193 in fully automated runs; the maximum score for a run in manual mode was 190.

4.2.3. *Experimental conditions*: Only the properties of the automatic pumps were manipulated; in manual mode the pumps were always accurate. This allowed a clearer interpretation of the effects of the properties of the *automation* on trust and intervention, free from any influence of the properties of manual control. The properties of the automatic pumps were manipulated on two dimensions—magnitude of control error and variability of control error—as summarized in table 4. Five magnitudes of control error spanned the range between a very good automatic pump for this system and a very poor one: depending on the experimental condition, the automatic pump would set the pump target to match the inflow (+0 condition), or set it to 5, 10, 20 or 40 litres greater than the inflow. In addition, two magnitudes of variability were added to the automatic pumps' responses: either zero variability (yielding a consistent control error), or a small Gaussian disturbance (0 mean, s.d. = 1, yielding a variable control error). To improve the transparency of the properties of the automatic pumps, the magnitude of control error and standard deviation of the Gaussian disturbance were constants in this experiment, compared to constant proportions in experiment 1. The two factors were completely crossed, yielding ten experimental conditions. Each condition was randomly mapped onto a pasteurizer, so there were ten pasteurizers in the plant, each with a unique automatic pump. All other components were consistently accurate.

4.2.4. *Training procedure*: Operators were trained intensively on a training pasteurizer whose control properties were exact. The objective of the rigorous training procedure was to allow operators to develop an effective allocation strategy between automatic and manual pump mode, supervising the better automatic pumps, but being able to intervene effectively in manual pump mode if necessary, as would expert operators in real systems. A part-whole task method of training emphasized, in stages 1–4 respectively: introducing the system using a detailed training manual, observing fully automated system performance, practising manual pump control and

Table 4. Summary of the ten experimental conditions in experiment 2. C = experimental condition. Sample values of the automatic pumps' manipulated pump target given in response to an inflow of 30 litres/minute are shown in square brackets.

Magnitude of the automatic pump's control error (litres/minute)	Variability of the automatic pump's control error	
	Consistent	Variable
+0	C1 [30·0]	C2 [29·7]
+5	C3 [35·0]	C4 [35·8]
+10	C5 [40·0]	C6 [38·4]
+20	C7 [50·0]	C8 [49·9]
+40	C9 [70·0]	C10 [71·2]

practising switching between modes. In stage 5, operators performed test exercises using the manual pump to produce and recover from specified unsafe vat volume states, as well as practising running the system in fully automatic mode, until their scores had reached asymptote. Operators were trained individually by the experimenter for one 90-min session per day, each (except for session 1) consisting of 10 runs, with each run lasting 100 iterations, about 8 min in real time.

4.2.5. *Subjective rating scales*: Operators were introduced to the trust rating scale at the beginning of their second session in training. This scale was 100 mm wide, and had 5 equally-spaced points with adjectival levels ('none at all, quite low, moderate, quite high, extremely high') to assist raters in using the middle part of the scale. Operators also rated their confidence in their rating of trust (i.e. how sure they felt about their trust rating). The confidence rating scale was identical to the trust rating scale. After a verbal introduction to trust in inanimate objects and several examples, operators practised using the scales by rating their trust and confidence in five everyday things (e.g. the public transit system to get you to your destination on time, your calculator to produce the correct answer). After every training run, operators rated their trust in the automatic pump and their confidence in their trust rating.

4.2.6. *Plant Preview*: Operators were not informed of the experimental manipulation, but to introduce them to the fact that the automatic pumps in the plant varied in quality and to help them use the scales effectively, they were given a plant preview which demonstrated the range of behaviour to be expected in the automatic pumps in the plant. Operators were explicitly informed of the relative quality of the automatic pump they were about to observe, then observed the training pasteurizer behaving like 'the best automatic pump in the plant', like 'an intermediate quality automatic pump', and like 'the worst automatic pump in the plant', for one run each. They could not intervene during the previews, but could monitor the pump rate as often as they wished.

4.2.7. *Experimental procedure*: Each operator worked for one session on each of the ten plant pasteurizers, in random order, controlling each pasteurizer for 8 runs of 100 iterations per run, and making subjective ratings of trust and confidence after each run.

4.3. *Results*
4.3.1. *Training*
4.3.1.1. *Scores*: All operators achieved asymptotic levels of performance during session 7, as determined by visual inspection of their learning curves.Thus, each operator received seven sessions (10·5 hrs) of training, of which 7·8 hrs was real time controlling the simulation –2·5 hrs in the first three teaching sessions and 5·3 hrs performing the test exercises. Creating and recovering from emergency scenarios during the test exercises forced operators' scores to be relatively low. Scores tended to improve over test exercise sessions (means = 137·3, 147·2, 151·5, and 153·5, respectively), $F(3, 15) = 2{\cdot}23$, $0{\cdot}05 < p < 0{\cdot}10$, but more important, the variance in scores decreased from a standard deviation of 22·0 in session 4 to 2·1 in session 7, t-test for equality of variances $(4) = 13{\cdot}66$, $p < 0{\cdot}001$. The significance of this is the operators' skills in using the automatic and manual pump had reached

asymptote, so whatever allocation strategy they chose in the experimental conditions could not be attributed to lack of manual skill.

4.3.1.2. *Subjective ratings of trust in the automatic pump*: The measure of trust was defined as the number of mm from 0 at the left pole, yielding scores from 0–100. Novice operators' trust grew along a typical learning curve during training, from a mean of 77·0 on the first run to 86·8 on the last run, $t(5) = 2{\cdot}14$, $p < 0{\cdot}05$, l-tailed. All operators' trust ratings had reached asymptote by session 5.

4.3.2. *Experimental conditions*

4.3.2.1. *Trust in the automatic pumps*: Operators made 8 trust/confidence ratings in each condition. On the assumption that real-life, expert operators would show high confidence in their trust judgements, those runs showing the highest confidence ratings were selected as the best estimate of 'expert' operators' trust in this task. Since confidence ratings were without exception highest on run 8, the ratings of trust on run 8 were used as a measure of experts' trust in the 10 automatic pumps in the plant.

Expert operators' mean ratings of trust in the automatic pumps are shown in figure 5. An analysis of variance confirmed that the magnitude of control error reliably influenced trust, $F(4, 20) = 31{\cdot}99, p < 0{\cdot}001$; the nature of this decline in trust is analyzed in more detail below. The main effect of variability was not significant, however, a marginally significant magnitude X variability interaction, $F(4, 20) = 2{\cdot}58$, $0{\cdot}05 < p < 0{\cdot}10$, suggested that variability may have influenced trust at some levels of the magnitude factor. A series of planned comparisons indicated that variability in the automatic pump's control error decreased trust for the +0 condition, $t(5) = 2{\cdot}73$, $p < 0{\cdot}02$, and tended to decrease trust in the +5 condition, $t(5) = 1{\cdot}92$, $p < 0{\cdot}06$, but did not affect trust in the +10, +20, and +40 conditions.

To characterize the nature of the decline in trust as a function of the magnitude of

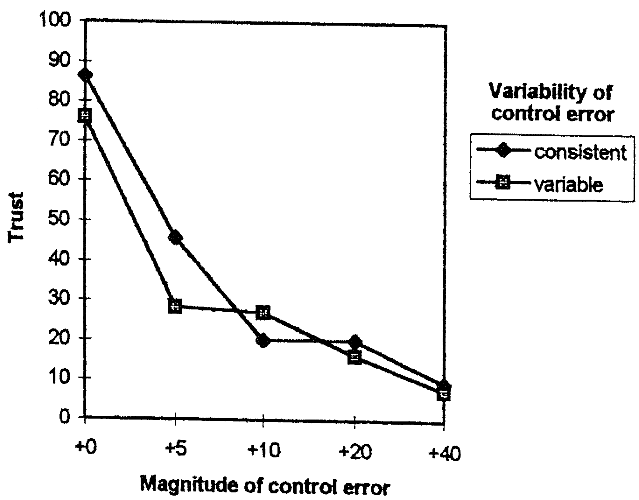

Figure 5. Trust and the properties of the automation. Expert operators' mean ratings of trust in the automatic pumps, as a function of the magnitude and variability of the automatic pumps' control error.

control error in the automatic pumps, regression analyses were performed. Separate analyses were performed for the consistent and variable error conditions since the variability factor tended to interact with the magnitude factor. A power function, as well as linear, quadratic, and log magnitude models were fitted and accounted for 34%, 45%, 66% and 72% of the variance in trust respectively in the consistent error conditions and 35%, 39%, 55% and 63% in the variable error conditions. Therefore, the log magnitude model was selected as the best model. The raw data points and fitted regression functions are shown in figure 6. For the consistent automatic pumps, the function was:

$$\text{Trust} = 83{\cdot}0 - 49{\cdot}0 \log_{10}(\text{Magnitude of control error} + 1)$$

and for the variable automatic pumps, the function was:

$$\text{Trust} = 70{\cdot}9 - 42{\cdot}0 \log_{10}(\text{Magnitude of control error} + 1)$$

Trust was greatly reduced by small control errors, but was increasingly less sensitive to larger control errors.

The development of operators' trust was examined by comparing their trust as novices (after the first run on each system) and as relative experts (after the last run on each system). Visual inspection of the data (figure 7) suggests that operators were initially conservative in their trust and adjusted their trust slightly with experience toward the extremes, according to the quality of the automatic pump, with trust increasing in the two best automatic pumps (Conditions 1 and 2) and decreasing in the poorer ones (Conditions 4–10). However, these changes are small and a series of planned comparisons indicated that only the change in trust in Condition 2 approached significance, $t(5) = 2{\cdot}04$, $0{\cdot}05 < p < 0{\cdot}10$. When data were pooled post hoc from Conditions 1 and 2 to represent 'trusted automatic pumps' and from Conditions 4–10 to represent 'distrusted automatic pumps' (the static Condition 3 data were not included), trust in the trusted automatic pumps increased with experience, $t(5) = 2{\cdot}84$, $p < 0{\cdot}05$, but the decline in trust for distrusted automatic pumps remained nonsignificant.

The trust ratings were also analyzed to determine whether operators' distrust of the poor automatic pumps in the plant generalized to reduce their trust in the best automatic pump (in Condition 1). The Condition 1 pasteurizer and the training pasteurizer were identical, both having exact automatic pumps, the only difference being that Condition 1 was embedded in the experimental sessions. Thus, any difference in trust between the two systems can be attributed to generalization of distrust in the poor automatic pumps to the best automatic pump. On the final run on the training pasteurizer, the mean rating of trust was 86·8. There was no difference between this and novice operators' trust in the automatic pump after run 1 in Condition 1 (81·3), $t(5) = 1{\cdot}00$, n.s., or expert operators' trust after run 8 in Condition 1 (86·3), $t(5) = 0{\cdot}13$, n.s.

4.3.2.2. *Use of the automatic pumps*: Operators' use of the automatic pump in each system had stabilized after four runs, so the mean of runs 5–8 was used as a best estimate of 'expert' allocation behaviour. Figure 8 shows the mean proportion of iterations spent in automatic pump mode. Operators' use of the automatic pumps was clearly affected by the magnitude of their control error, $F(4, 20) = 34{\cdot}5$, $p < 0{\cdot}001$ and marginally by the variability factor, $F(1, 5) = 6{\cdot}0$, $0{\cdot}05 < p < 0{\cdot}10$. The two factors did not interact. With respect to the magnitude of error main effect, post hoc

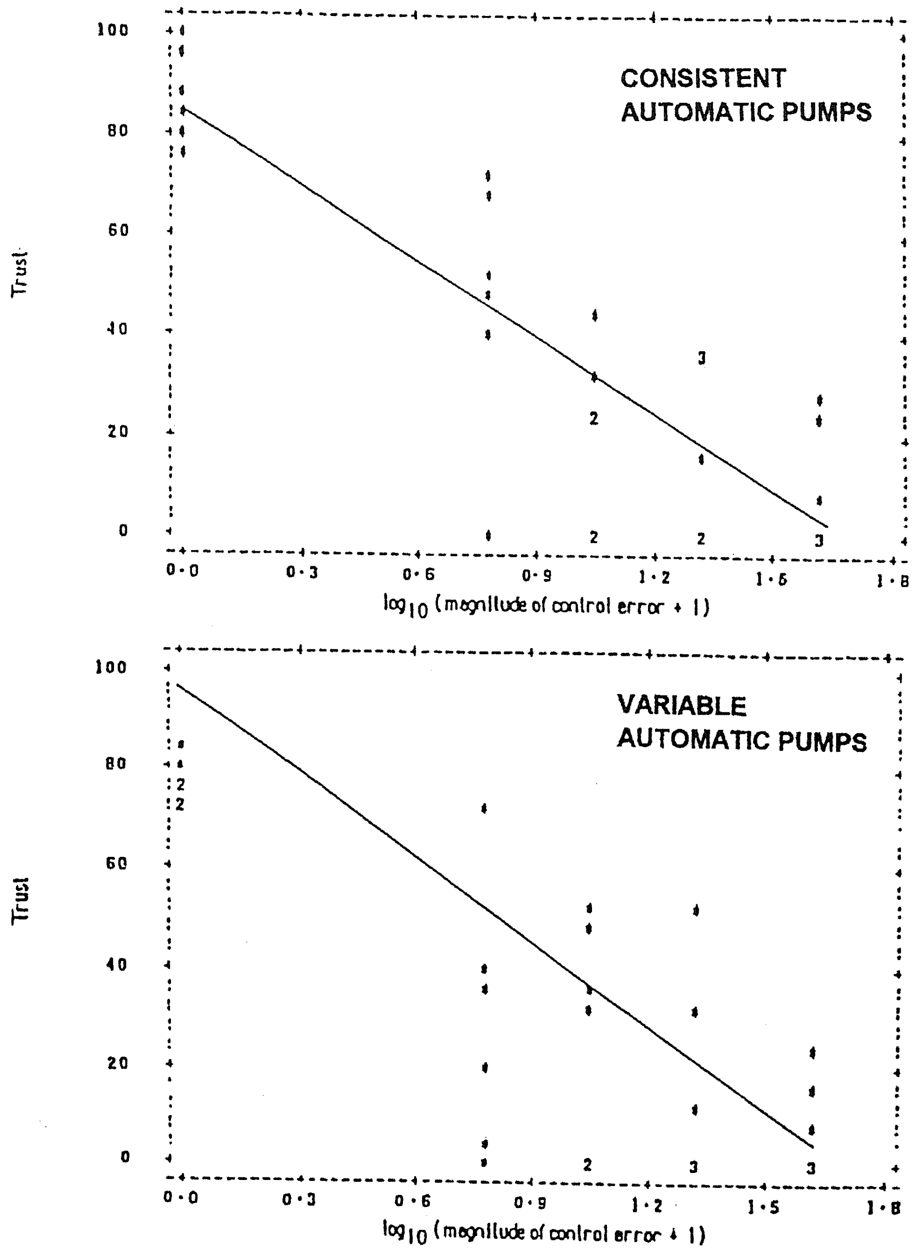

Figure 6. The relationship between trust and the properties of the automation. Raw data points and fitted regression function of expert operators' trust on the automatic pumps' magnitude of control error, for consistent and variable automatic pumps.

multiple comparisons (Newman-Keuls tests at $a = 0{\cdot}05$) revealed that for the consistent automatic pumps, operators used the +0 automatic pump more than the +5 and use of the +5, +10, +20, and +40 automatic pumps did not differ. The pattern of results was the same for the five variable automatic pumps. A series of planned

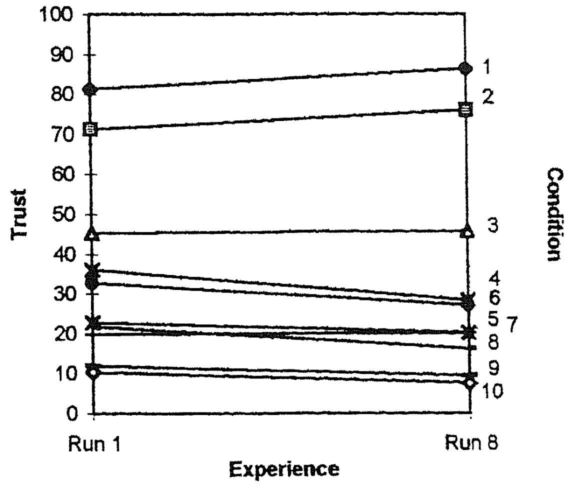

Figure 7. Changes in trust with experience. Mean ratings of novice (run 1) and expert (run 8) operators' trust in the automatic pumps in the ten experimental conditions.

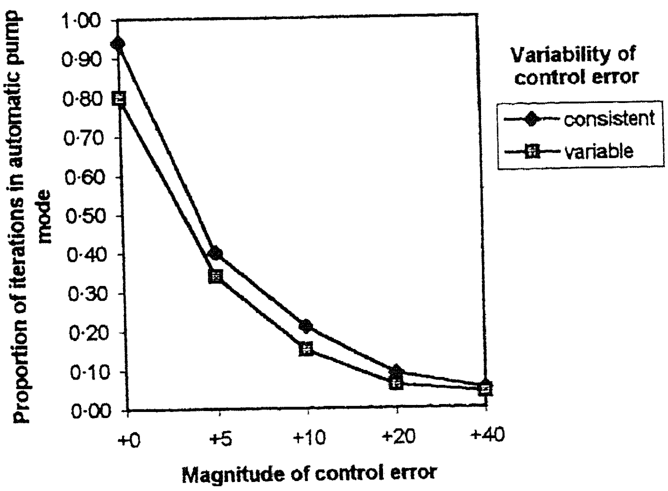

Figure 8. Use and the properties of the automation. Mean proportion of time expert operators spent in automatic pump mode, as a function of the magnitude and variability of the automatic pumps' control error.

comparisons between the consistent and variable automatic pumps revealed that the effect of variability was significant only at the +0 magnitude of control error, $t(5) = 2{\cdot}58$, $p < 0{\cdot}05$.

4.3.2.3. *Relationship between trust and use of the automatic pumps*: The relationship between operators' subjective ratings of trust and their intervention behaviour is of primary importance in this experiment. A correlation was computed between the measures of 'expert' (as defined above) operators' trust and use, pooled over all conditions. There was a high positive correlation between operators' trust in the automatic pumps and the time they spent in automatic pump mode, both pooled over operators, $r(58) = 0{\cdot}71$, $p < 0{\cdot}0001$ and for all individual operators (see table 5). The nature of this relationship was examined more closely with regression analysis. The fitted linear regression is shown in figure 9. The measure of use (the proportion of iterations in automatic pump mode) was multiplied by 100 to yield a more interpretable regression equation in terms of the predicted number of iterations of automatic pump use per 100-iteration run:

$$\text{Number of iterations in automatic pump mode} = 3 + 0{\cdot}8(\text{Trust})$$

Table 5. Correlation between operators' trust in the automatic pumps and the time spent in automatic pump mode, for the group of operators and each individual operator, pooled over the ten experimental conditions.

Operator	Correlation Coefficient	df	p
Pooled	0·71	58	< 0·0001
1	0·83	8	< 0·005
2	0·75	8	< 0·02
3	0·87	8	< 0·001
4	0·94	8	< 0·001
5	0·71	8	< 0·05
6	0·93	8	< 0·001

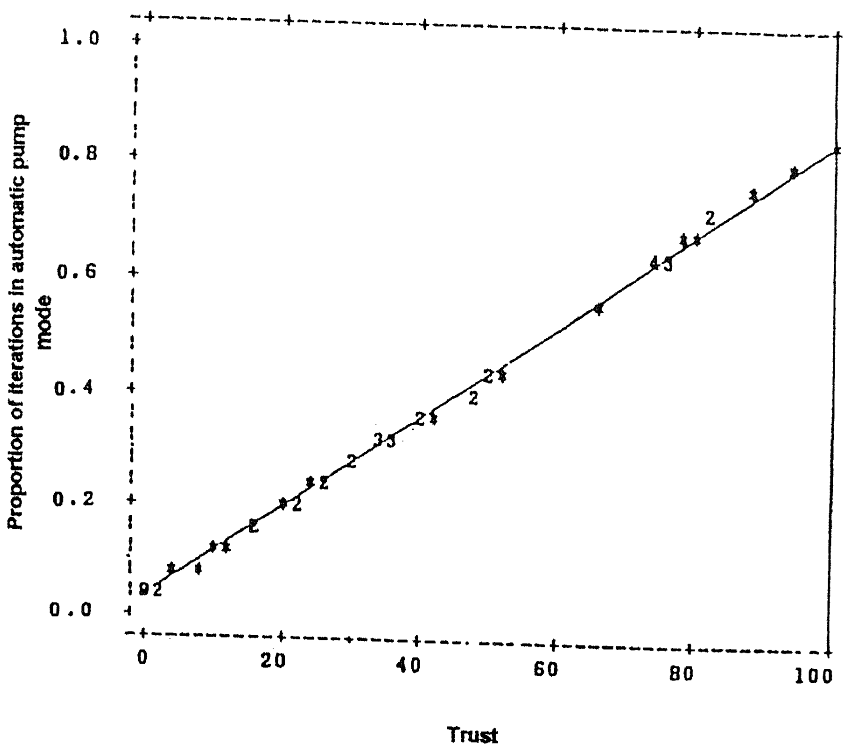

Figure 9. The relationship between trust in and use of the automation. Fitted linear regression of expert operators' use of the automatic pumps on trust in the automatic pumps.

This regression model was highly reliable, $F(1, 58) = 60 \cdot 17, p < 0 \cdot 0001$. The intercept value failed to reach statistical significance, $t(58) < 1$; the 95% confidence interval for the intercept was from $-0 \cdot 07$ to $0 \cdot 13$ (on the original proportional scale). The slope of the regression was highly reliable, $t(58) = 7 \cdot 76$, $p < 0 \cdot 0001$; the 95% confidence interval for the slope was between 0·6 and 1·0 iteration of use per unit increase in trust. Thus, the relation between use of the automatic pumps and trust approached 1:1, falling short because use was more conservative than trust. This relation held, not only pooled over operators, but also for every individual operator (see table 6). The regression equations are remarkably similar for all six operators, the F values are all significant and the R^2's are high.

4.3.2.4. *Relationship between monitoring and trust*: The number of times that operators checked the automatic pumps' pump rate had stabilized after four runs, so the mean of runs 5–8 in each experimental condition was used as a best estimate of 'expert' monitoring behaviour. It was necessary to correct these data to eliminate an artifact of the task demands, namely that operators had to be in automatic pump mode to check the automatic pump's pump rate. Since operators spent less time in automatic pump mode with poor automatic pumps, they had less opportunity to check the poor automatic pumps. The monitoring measure was equated for opportunity by calculating the number of checks of each automatic pump per iteration in automatic pump mode. In the +0, +5, +10, +20 and +40 control error conditions respectively, the mean number of checks of the automatic pump per iteration in automatic pump mode were 0·11, 0·35, 0·26, 0·60 and 0·66 for the consistent error pumps and 0·32, 0·46, 0·52, 0·73 and 0·68 for the variable error pumps. Monitoring was influenced by the magnitude of the control error, $F(4, 20) = 4 \cdot 09$, $p < 0 \cdot 025$; the poorer the automatic pump, the more operators checked it. Monitoring was also affected by the variability manipulation, $F(1, 5) = 33 \cdot 0$, $p < 0 \cdot 005$, with operators checking the variable automatic pumps more than the consistent ones. The two factors did not interact. As predicted, pooled over conditions and operators, there was a negative correlation, $r(58) = -0 \cdot 41$, $p < 0 \cdot 005$, 1-tailed, between expert operators' monitoring behaviour and their ratings of trust: the less operators trusted an automatic pump, the more intensely they monitored it and the more they trusted an automatic pump, the less intensely they monitored it.

4.3.2.5. *Scores*: Performance scores had stabilized after four runs, so the mean of

Table 6. Summary of regression analyses of expert operators' use of the ten automatic pumps on their trust. 'Use' is the predicted number of iterations in automatic pump mode per 100-iteration run.

Operator	Regression Equation	R^2 Model	F(Reg)	p
Pooled	Use = 3 + 0·8 Trust	0·51	60·17	< 0·0001
1	Use = 27 + 0·8 Trust	0·68	17·08	< 0·005
2	Use = −7 + 1·4 Trust	0·56	10·10	< 0·025
3	Use = −9 + 1·0 Trust	0·76	25·43	< 0·001
4	Use = 4 + 1·0 Trust	0·89	64·64	< 0·0001
5	Use = −17 + 0·8 Trust	0·51	8·32	< 0·025
6	Use = −23 + 1·2 Trust	0·86	49·50	< 0·001

runs 5–8 was used as a best estimate of 'expert' performance on each plant system. The penalty for use of the manual pump was included in the scoring system to discourage operators from competing indiscriminately with the automatic pump. The data on use of the automatic pump show that it accomplished this purpose. However, the penalty artifactually reduced scores in the conditions with poorer pumps where manual intervention was appropriate. Therefore, it was eliminated post hoc from the scoring system to provide a measure based on productivity and safety, independent of the pump mode used to achieve this score. In the +0, +5, +10, +20, +40 control error conditions respectively, expert operators' mean corrected scores were 191·0, 190·0, 189·4, 187·4, 188·3 for the consistent automatic pumps and 191·1, 190·3, 190·8, 189·3, 190·5 for the variable automatic pumps. Scores were not affected by the magnitude or the variability of the automatic pumps' control error, both Fs< 1. Operators successfully controlled all ten systems, adapting their allocation strategy as necessary, and expertly executing manual control when necessary, to maximize their scores.

4.3.2.6. *Relationship between scores and trust*: Experts' scores did not correlate significantly with their trust, $r(58) = 0{\cdot}04$, n.s., indicating that operators did not base their trust on performance.

4.4 *Discussion*

The most important finding of experiment 2 was the high positive correlation between operators' trust in the automatic pumps and the time they spent in automatic pump mode. This correlation, highly reliable for the group of operators studied as well as for each individual operator, indicates that operators' subjective ratings of trust provide a simple, nonintrusive insight into their use of the automation. Regression analyses indicated that the relation between trust and use approached 1:1. The particular intercept and slope are no doubt specific to this task, but the clear direction of the association provides strong support for the hypothesis (Sheridan and Hennessy 1984) that operators will use the automation only to the extent that they trust it; if they do not trust the automation, they will reject it, preferring to do the task themselves manually.

The regression of use on trust in this study is consistent with conventional wisdom in the field, with case study results (Zuboff 1988) and with our own intuition in everyday situations, that trust *determines* people's use of automation. However, the present results demonstrate only a correlation between trust and use and cannot distinguish whether the properties of the automation determined trust, which in turn determined use, or whether the properties of the automation concomitantly affected both trust and use. An extension of this research (Lee and Moray 1992), using a similar simulation task, has established that trust is indeed a causal variable, mediating between the properties of the automation and the operator's allocation behaviour. This has important implications for retraining operators whose trust and/or use of the automation is found to be systematically biased. Since trust determines intervention behaviour, recalibration training should focus on modifying operators' trust in the automation and use should follow. Several methods for recalibrating trust have been suggested by Muir (1987, 1994), for example, improving the accuracy of operators' perceptions of machine competence, a recommendation that now has empirical support.

The results of experiment 1 showed that competence is an important determinant of trust in the automation. The results of experiment 2 confirmed and extended this finding: both of the manipulations of the automatic pumps' competence—the magnitude and the variability of the control error—influenced operators' trust in the automation. As predicted, operators trusted the exact automatic pump and their trust diminished as the magnitude of the control error increased. But the decline was not a gradual one; rather, trust fell precipitously from the +0 to the +5 condition and less steeply from there. Operators' moderate trust in the +5 magnitude condition coincides with the system's tolerance for pump rate changes of about 5 litres/min. The +5 error would sometimes (depending on other system variables) momentarily sacrifice system performance and this is reflected in operators' lowered trust; the +10 and larger magnitudes of control error consistently impaired system performance and operators distrusted them even more. This finding replicates that of experiment 1 in which a control error's consequences for system control appeared to be an important determinant of trust. It also extends this finding to show that the consequences need not be *consistently* negative to diminish trust; trust is fragile and any instances of even *temporary* loss of control are sufficient to reduce it.

As predicted, a small degree of variability in the control error reduced operators' trust, but only when variability was added to an exact automatic pump; otherwise, the variability factor was washed out by the more potent magnitude of control error factor. Over the course of fully automated runs, the added variability did not degrade system performance; however, from moment to moment it sometimes actually improved the automatic pumps' effectiveness, sometimes reduced it slightly and usually had no effect. That variability reduced operators' trust under these circumstances suggests that operators based their trust not on long-term, usual or positive consequences for control of a variable automatic pump, but rather on immediate, specific, negative consequence for control. This finding confirms that operators' trust will be reduced not only if a control error always degrades system performance, but also if it does so only momentarily, even if this does not affect overall system performance. Operators' trust is apparently conditionalized on the worst observed machine behaviours. This is a desirable strategy in real supervisory control systems, which are usually capital intensive and dangerous. If there is any variability in the quality of an automated subsystem's output, operators cannot afford to base their trust on its long-term average performance or its short-term usual or best performance because it is a subsystem's worst performance that will threaten productivity and safety.

Taken together, the effects of the magnitude and variability manipulations are consistent with the hypothesis (Sheridan and Hennessy 1984, Rempel *et al.* 1985) that behaviour must be both desirable and consistent to foster trust. However, this conclusion can now be further qualified on the basis of the present results. The most desirable and consistent automation was indeed the most trusted in this study and reducing either of these qualities slightly did diminish trust substantially—but not to zero. Operators' trust is not an all-or-nothing variable, existing only with necessary conditions. Rather, trust is a continuous variable and can remain relatively high in the face of small degrees of undesirability or inconsistency.

Operators' trust changed very little with experience in this experiment. In general, operators anchored on their initial ratings of trust in each system. Although there was a consistent trend for operators' ratings to become more polarized with experience, this effect was statistically significant only for the growth of trust in 'trusted automation', as defined post hoc. This finding supports the hypothesis (Muir 1994) that distrust is more resistant to change than trust. An important implication of this inertia, emphasized also by Zuboff (1988), is that care must be taken in the introduction of new automation because operators' trust, whether appropriate or not, may persist at initial levels.

The results of this experiment indicated that operators' distrust of poor automatic pumps did not spread across independent, but highly similar, systems to diminish their trust in the best automatic pump in the plant. Of course, the converse, though not usually articulated, is also true: neither did operators' trust in the best automatic pump generalize to prevent distrust of the poorer ones. Operators discriminated between systems, conditionalizing their trust on the particular properties of individual systems. Data from questionnaires completed after each session (reported in Muir 1989) indicated that operators had fairly accurate perceptions of the automation's properties, suggesting that the present results may bear some resemblance to the behaviour of experienced operators in the field, who also have fairly accurate perceptions of the properties or 'signatures' of systems (Woods 1986).

The results of this experiment supported the predicted inverse relationship between operators' trust and monitoring of the automation. This finding provides empirical support for Sheridan and Hennessy's (1984) hypothesis that supervisors spend more time monitoring the automation they distrust and also for Rempel *et al.*'s (1985) hypothesis that people rely less on observations of behavioural acts of trusted referents and more on their own attributions about the referent. The low level of monitoring of trusted automation in this experiment parallels real-world operators' tendency to become complacent in their monitoring of good automation (e.g. Wiener and Curry 1980). It is important to note that this inverse relationship should emerge only when distrusted automation *must* be used, as it had to be in this experiment, at the beginning of runs, when operators had to discover the properties of the automation before making an allocation decision, or as it would have to be in automated systems which lack or discourage manual override. However, if supervisors can override distrusted automation, from that point, the inverted U-shaped function predicted by Muir (1994) should emerge, with distrusted automation being monitored infrequently, it at all, because it is out of the control loop.

5. Summary and conclusions

5.1. *The integrated model of trust in machines*

The first dimension of the integrated model of trust in machines proposed by Muir (1994) provided a theoretical basis for hypothesizing that the competence of the automation is an important aspect of human-machine trust. The results of the two experiments reported here confirmed this and therefore validate the model. While it is useful to know that competence is an important determinant of trust in a machine, it is necessary to examine the expectation of competence more closely if we are to develop a predictive model of trust in machines and intervention behaviour. The two experiments reported here examined the effect of different kinds of chronic incompetencies on operators' trust. Experiment 1 examined the importance of two qualitative aspects of automation—its control and display properties—and showed that both of these aspects of machine competence affect trust in an additive way. This qualitative analysis was extended in experiment 2 to a quantitative analysis of the effect of different-sized control errors on trust, to investigate the tolerance limits of trust. Results showed that operators' trust was eroded substantially by small consistent control errors and reached asymptote at the point where the control error was large enough to consistently compromise system performance. Trust was also reduced if the automation's responses were variable, even though the degree of disturbance was too small to affect overall system performance. Taken together, the results of both studies suggest that operators' trust is sensitive to the properties of the automation and is easily eroded by any indication of incompetence which impairs system performance, even if the impairment is only temporary and of no consequence to long-term system performance. However, trust in minimally incompetent automation was not shattered, falling to zero; operators expressed moderate degrees of trust in these cases, treating trust as a quantitative, rather than as a qualitative, all-or-none variable. Trust is fragile, but not brittle. Now that an empirical foundation has been laid, more research is needed to investigate the role of operators' expectations of persistence and machine responsibility in determining trust.

Tests of the second dimension of the integrated model of human-machine trust, the development of trust with experience, showed that the constructs of predictability, dependability and faith are important aspects of developing trust, but more research is needed to determine their relative contribution at different times during the human-machine interaction. The developmental dimension of the model suggested important hypotheses about the factors which affect the growth or erosion of human-machine trust, for example, the observability, transparency and consistency of machine behaviours, as well as suggesting the psychological attribution process as the mechanism underlying the differential monitoring of trusted and distrusted automation.

In conclusion, the integrated model of trust in machines (Muir 1994) provided an organized theoretical foundation for research in this area and suggested a rich variety of testable hypotheses with may not have been formulated otherwise. Many of these hypotheses received empirical support in the experimental data reported here suggesting that this model should be retained to guide future research. It has already been used as a springboard for other research to extend the present findings. For example, Lee and Moray (1992) have gone on to examine the destruction and recovery of trust in the presence of acute and chronic faults, choosing their response variables and interpreting their results in terms of Barber's (1983) and Rempel *et al.*'s (1985) constructs. They have also investigated an individual difference—the role of operators' self-confidence in their manual control skill—as a determinant of intervention behaviour. This factor was controlled in the present experiments by the very intensive training of operators in both manual and automatic control modes and by the setup of the experimental task which encouraged supervisory control whenever possible, rather than a reliance on manual control.

5.2. *The calibration of trust*

The appropriateness of 'calibration' of trust is an important issue in automated systems and is discussed in more detail in Muir (1987) and in Part I of this article (Muir 1994). If operators' trust in the automation is systematically biased with respect to objective measures of its performance, then recalibration training should improve the calibration of their trust and their use of the automation. The underlying assumption is that trust *can* be changed, that it is responsive to changes in the operator's perception of system properties. Data from the two experiments reported here suggest that this assumption is warranted. Although, in general, operators adjusted their trust very little with experience, there were two interesting exceptions. In experiment 1, operators' trust in the pumps with a constant control error increased with experience. In experiment 2, trust in the two best automatic pumps grew when these were combined to represent 'trusted automatic pumps.' Taken together, these results show that operators can improve the calibration of their trust of their own accord, so we can be optimistic that retraining which improves the accuracy of their perception of system properties will also improve the appropriateness of their trust in the automation. The pattern of circumstances under which operators adjusted their trust supports the following conclusions about the growth of trust in automation:

- *The automation must be in the control loop for trust to grow.* In experiment 1, operators could not escape the automatic pumps' errors because the manual pumps had the same errors. They had to learn to compensate for the pumps' errors to maximize their score and their trust in the constant error pumps grew. In experiment 2, only the automatic pumps' properties were manipulated and operators could escape the error by using manual pump mode. Consequently, they did not have to find a strategy to work with the error in the poor automatic pumps, other than simply taking the automatic pump off-line. Trust in these automatic pumps was low and did not change with experience. The only automatic pumps for which trust grew were the two best ones, which operators used extensively. This pattern of results provides empirical support for two hypotheses developed in Part I of this article (Muir 1994), that behaviours must be observable for trust to grow and that distrust is more resistant to change than trust. If distrusted automation is taken off line, its behaviours are no longer observable and there exists no opportunity for operators to reassess their perceptions of it or their trust. An implication of this conclusion is that we cannot expect operators' trust in a specific part of an automated system to change as a result of experience on a *system*; experience with the specific automation is a necessary condition for trust to grow.
- *Having the automation in the control loop is a necessary, but not a sufficient, condition for the growth of trust.* Although in experiment 1, operators could escape from neither the constant nor the variable control errors, their trust grew only in the constant error pumps.
- *Trust will grow if operators can find a strategy to compensate for the consequences of an error in on-line automation.* In experiment 1, operators discovered the rule that governed the constant control error pumps, they learned to compensate for it in their manual control and their trust grew. In contrast, there was no way to compensate for the variable control error pumps and trust in these did not grow.

5.3. *Trust—a necessary and valuable subjective variable*

Operators' performance scores were high in both of the present experiments and were unaffected by all of the experimental manipulations of the properties of the automation. Were one to look only at this measure of performance, one would conclude that the properties of the automation had no effect on the operator or on system performance. However, operators' subjective ratings of trust showed large and orderly effects of the manipulations and furthermore, were strongly associated with *how* the operators performed their task—whether they chose to use the automation, or reject it and perform the task manually.

This pattern of results underscores the necessity of understanding the operator's subjective relationship with the automation; we cannot be satisfied simply with an acceptable level of system performance, without understanding how this is being achieved and at what personal price for the operator. For example, consider two operators, both of whom maintain satisfactory system performance: one trusts the automation, uses it and monitors it intermittently while also performing other higher-level tasks such as planning. The other operator distrusts the same automation, intervenes frequently and monitors it intensively to the detriment of higher-level tasks. Short-term system output may be identical in both cases, but the operators' personal experiences are quite different. The operator who distrusts the automation will have a poorer quality of working life, perhaps experiencing alienation (Sheridan 1980), increased levels of stress, time pressure and mental workload, which may render system performance brittle and vulnerable to failure. The results presented here show that operators' subjective ratings of trust can provide a simple, nonintrusive insight into their personal relationship with the automation and this can be used as a basis for predicting and optimizing operators' allocation behaviour and quality of working life, as well as long-term system performance.

5.4. *Applications*

The results of these experiments can be applied usefully by the following people involved in automated systems:

- *Designers*, in deciding whether to automate certain functions, should consider carefully whether the function can be automated very effectively. If operators detect any weakness in the automation, they will not trust it and given the choice, will not use it. Consequently, systems should be designed to support operators in manual control mode as well as in automatic mode.
- *Training personnel* should carefully introduce novices to the automation or experts to new automation, to optimize the calibration of their initial trust. If the need arises, it may be possible to improve the calibration of operators' trust (see also Part I of this paper, Muir 1994, Muir 1987).
- *Management*, by recognizing the importance of the operator's subjective relationship to the automation and allocating resources to understand and improve it, can optimize both the quality of the operator's working life and system performance.
- *Operators*, equipped with knowledge and understanding of their own attitudes toward and use of the automation, can facilitate the calibration process, thereby improving their quality of working life and system performance.
- *Researchers* can use the theoretical foundation and empirical results presented here to do the field studies which are necessary to validate these laboratory results and further our understanding of trust and human intervention in a variety of automated systems.

Acknowledgements

The author wishes to thank Neville Moray for his contributions as supervisor of this doctoral research and two anonymous reviewers for their helpful comments on an earlier draft. Financial support was provided by the Natural Sciences and Engineering Research Council of Canada. This work was performed in the Department of Industrial Engineering, University of Toronto, Canada.

References

BARBER, B., 1983, *The Logic and Limits of Trust* (Rutgers University Press, New Brunswick).

GOLDBERG, L. R., 1968, Simple models or simple processes? Some research on clinical judgments, *American Psychologist*, **23**, 483–496.

LEE, J., 1992, Trust, self confidence, and operators' adaptation to automation, Doctoral dissertation. University of Illinois at Urbana-Champaign.

LEE, J. and MORAY, N., 1992, Trust, control strategies and allocation of function in human-machine systems, *Ergonomics*, **35**, 1243–1270.

MUIR, B. M., 1987 Trust between humans and machines, and the design of decision aids, *International Journal of Man-machine Studies*, **27**, 527–539. Also in E. HOLLNAGEL, G. MANCINI and D. D. WOODS (eds), 1988, *Cognitive Engineering in Complex Dynamic Worlds* (Academic, London).

MUIR, B. M., 1989, Operators' trust in and use of automatic controllers in a supervisory process control task, Doctoral dissertation, University of Toronto, Canada.

MUIR, B. M., 1994, Trust in automation. Part I. Theoretical issues in the study of trust and human intervention in automated systems, *Ergonomics*, **37**, 1905–1922.

REMPEL, J. K. and HOLMES, J. G., 1986, How do I trust thee? *Psychology Today*, February, 28–34.

REMPEL J. K. and HOLMES, J. G. and ZANNA, M. P., 1985, Trust in close relationships, *Journal of Personality and Social Psychology*, **49**, 95–112.

SHERIDAN, T. B., 1980, Computer control and human alienation, *Technology Review*, October, 61–73.

SHERIDAN, T. B. and HENNESSY, R. T. (eds), 1984, *Research and Modeling of Supervisory Control Behavior* (National Academy Press, Washington).

WIENER, E. L. and CURRY, R. E., 1980, Flight-deck automation: promises and problems, *Ergonomics*, **23**, 995–1011.

WOODS, D. D., 1986, The design of decision aids in the age of 'intelligence,' In *Proceedings of the 1986 IEEE International Conference on Systems, Man, and Cybernetics*, Atlanta, 398–401.

WOODS, D. D., ROTH, E. M. and BENNETT, K., 1987, Explorations in joint human-machine cognitive systems, W. Zachary and S. Robertson (eds), *Cognition, Computing and Cooperation* (Ablex, Norwood).

ZUBOFF, S. 1988, *In the Age of the Smart Machine: The Future of Work and Power* (Basic, New York).