Minimal Instrumentation May Compromise Failure Diagnosis With an Ecological Interface

Dal Vernon C. Reising, University of Illinois at Urbana-Champaign, Urbana, Illinois, and **Penelope M. Sanderson,** University of Queensland, Brisbane, Australia

Interfaces designed according to ecological interface design (EID) display higher-order relations and properties of a work domain so that adaptive operator problem solving can be better supported under unanticipated system conditions. Previous empirical studies of EID have assumed that the raw data required to derive and communicate higher-order information would be available and reliable. The present research examines the relative advantages of an EID interface over a conventional piping-and-instrumentation diagram (PID) when instrumentation is maximally or only minimally adequate. Results show an interaction between interface and the adequacy of the instrumentation. Failure diagnosis performance with the EID interface with maximally adequate instrumentation is best overall. Performance with the EID interface drops more drastically from maximally to minimally adequate instrumentation than does performance with the PID interface, to the point where the EID interface with minimally adequate instrumentation supports nonsignificantly worse performance than does the equivalent PID interface. Actual or potential applications of this research include design of instrumentation and displays for complex industrial processes.

INTRODUCTION

> What can be sensed forms a fundamental limiting feature of displays. This limiting feature is not always given the emphasis it deserves.
>
> –C. R. Kelley (1968, p. 90).

Ecological interface design (EID; Vicente, 2002; Vicente & Rasmussen, 1992) is an approach to designing the visual displays of human-machine interfaces that has grown in popularity in the cognitive engineering community over the last 15 years. An EID interface is intended to display the higher-order relations and properties of a work domain so that adaptive operator problem solving is well supported, particularly in unanticipated system conditions. EID principles suggest that such higher-order relations and properties should be derived from a work domain analysis of the system in question and that the information should be displayed so that operators can use direct perception and direct manipulation as much as possible when interacting with the system (Vicente & Rasmussen, 1992). Empirical research to date suggests that displays designed according to these principles do provide performance benefits, especially under unanticipated conditions, and some transfer of the ideas to industry has started. (For a summary see Vicente, 2002; for recent examples see Bisantz et al., 2003; Burns, 2000; Ham & Yoon, 2001a, 2001b; Jamieson, 2003; Jamieson & Vicente, 2001; and Yamaguchi & Tanabe, 2000.)

In one of the first presentations of EID, Vicente and Rasmussen (1992) expressed concern about the potential impact of sensors on interfaces designed according to EID principles. This concern has been repeated elsewhere (Beltracchi, 1998; Reising & Sanderson, 1996; Vicente et al., 1996). In a recent review, Vicente (2002) noted again the problem of sensor failure as an unresolved challenge for EID: “Many important issues have yet to be addressed, let alone solved.

Address correspondence to Dal Vernon Reising, Honeywell Laboratories, 3660 Technology Dr., Minneapolis, MN 55125; dalvernon.reising@honeywell.com. **HUMAN FACTORS,** Vol. 46, No. 1, Spring 2004, pp. 316–333.

Some of these issues, such as sensor failure, may turn out to be 'show stoppers'" (p. 75).

A first possible problem is that in order to support direct perception, EID interfaces often convey information through the space-time behavior of graphical forms. Therefore, the more that a display is configural (Bennett, Nagy, & Flach, 1997) – that is, the more the arrangement of parts adds information beyond that in the parts alone – the more devastating the impact of a faulty sensor might be. This is a concern that would extend to graphical interfaces that are not necessarily developed according to EID principles.

A second possible problem is that EID depends in part on being able to represent higher-order properties of system functioning. However, direct sensing of higher-order properties is often not available – sensors fail, sensors have not been installed, or no sensors exist to measure the property in question. (See Reising & Sanderson, 2002b, for a full discussion.) We have identified two forms of information inadequacy that result from sensor inadequacies: topographic and derivational. They are shown in Figure 1. Topographic inadequacy results when a physical reading from a sensor is borrowed to indicate the same type of reading (e.g., volume flow rate) at relatively remote locations in plant topology. Derivational inadequacy results when higher-order properties (e.g., mass flow rate) are derived from topographically inadequate lower-order information (e.g., volume flow rate). In such circumstances, it can be difficult to tell whether an apparently abnormal system state is a failure of the process or a failure of the sensors.

Reising and Sanderson (1996, 2002b, 2002c) developed a special kind of work domain analysis, called a *sensor abstraction hierarchy* (sensor AH), to identify information needs in ecological interfaces and to test for possible sensor inadequacies. The sensor AH superimposes a lattice-like inventory of sensors and measures over a lattice-like representation of the levels of the abstraction hierarchy. Reising and Sanderson (1996, 2002b, 2002c) used sensor AHs of a simulated pasteurization plant to demonstrate the relative robustness of a configural display for energy flow rate with two versions of an ecological interface. One version had enough sensors to measure temperature and volume flow on each leg of the plant separately and therefore independently derived mass and energy flow rate for each leg (maximally adequate). The other version had only the minimum number of sensors needed to derive the higher-order properties required by the ecological interface and therefore lacked any corroborating measurements elsewhere (minimally adequate). A cognitive

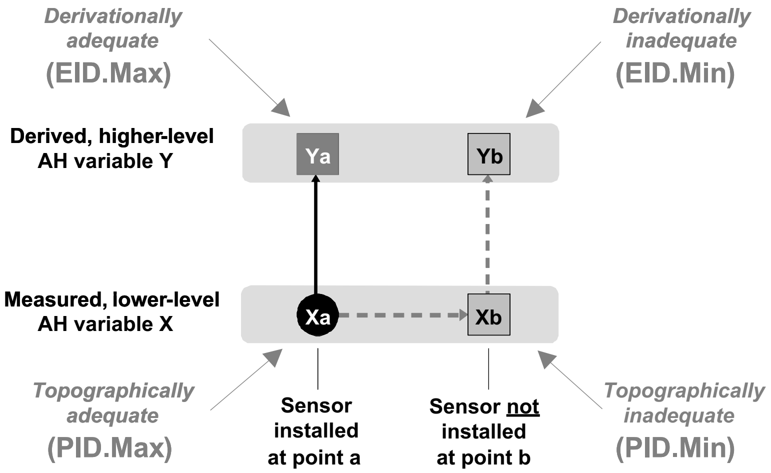

Figure 1. Relationship between topographically and derivationally adequate and inadequate information. Rows indicate two levels (higher, lower) of a sensor abstraction hierarchy (AH). EID.max, EID.min, PID.max, and PID.min are names of the interfaces (and experimental conditions) in which each kind of inadequacy tends to appear.

walkthrough of the reasoning involved in detecting and diagnosing system and sensor failures suggested that topographic and derivational inadequacy may make it difficult or even impossible to diagnose certain failures and that configural displays may sometimes add to the difficulty.

The adequacy of sensors has been examined both analytically and empirically in other work as well. The functional information profile (FIP) developed by Burns and Vicente (1996) provides an abstraction-decomposition framework for inventorying information to be conveyed in an interface: variables, targets, and boundary values. By implication, the FIP identifies the information that will need to be sensed. In later work, Jamieson (1998) used a similar approach to inventory the variables in a petroleum refining simulation that should be represented in an ecological interface. Although the representations by Burns and Vicente and by Jamieson (1998) included information at various levels of the abstraction hierarchy, the tabular formats do not make it easy for analysts to determine the effect of missing measurements or derivations on other variables. In contrast, the graphical lattice-based sensor AH used by Reising and Sanderson (1996, 2002b, 2002c) makes it easy to see how information specified in an FIP might be collected or derived.

The issue of building EID interfaces on an adequate sensor foundation is as important for system upgrades as it is for entirely new system development. Human factors engineers in the nuclear and petrochemical industries are increasingly required to help replace aging analog instrumentation and displays with advanced display technology. In some cases they are faced with legacy instrumentation that has serious shortcomings for displaying higher-level properties. Sometimes the sensors are minimally adequate to derive higher-order information, but sometimes the sensors are simply unavailable for such derivation. Efforts to apply the concepts of EID to industrial systems (Beltracchi, 1998; Hayter, 1996; Jamieson, 1998) have highlighted the need to address what effect the positioning and reliability of instrumentation might have on the visual forms of ecological interfaces and, by implication, on the performance of operators using ecological interfaces. The question then becomes a design trade-off among (a) what can be displayed without additional sensors, (b) what should be displayed at the cost of additional sensors, and (c) what can be left undisplayed.

In contrast, when a system is being developed de novo, an EID-motivated analysis of the raw information needed to display higher-order properties of a system can drive the requirements engineering process so that plant instrumentation is appropriate from the outset (Reising & Sanderson, 1996).

Our goal in this paper is to report the results of an empirical study that compared human performance for system and sensor failure diagnosis with an ecological and a nonecological interface, in which each interface was supported by either maximally or minimally adequate instrumentation. On the basis of demonstrations and walkthroughs in Reising and Sanderson (2002c), we anticipated that any interface – ecological or nonecological – would support better human performance for diagnostic tasks when the interface was supported by maximally as opposed to minimally adequate instrumentation. Moreover, on the basis of previous empirical research findings (see Vicente, 2002, for a summary), we predicted that an effectively designed ecological interface would support better human failure-diagnosis than a nonecological interface would.

However, the walkthroughs in Reising and Sanderson (2002c) suggested that ecological interfaces may be particularly vulnerable to minimally adequate instrumentation during abnormal process conditions as well as when instrumentation is unreliable or fails. Displays may become misleading if the display programming logic does not account for faulty process or instrumentation input (Leveson, 1995). The purpose of the research reported in this paper is to investigate the conditions under which ecological interfaces will safely deliver their benefits and to point to the kind of analyses needed to ensure those benefits.

In the sections that follow, first we describe the Pasteurizer II simulation used for the empirical study. Then we briefly describe the process by which an ecological interface was developed for Pasteurizer II and indicate how the distinction was made between the maximally and minimally adequate sensor configuration. Then we move to the methods and results of the empirical study that was performed to test our hypotheses. Finally, we discuss the implications of our findings for interface design using the EID approach.

Description of Pasteurizer II

The Pasteurizer II simulation has been described in detail elsewhere (Reising, 1999; Reising & Sanderson, 2002a, 2002c), so we provide only basic information here. Figure 2 is a piping-and-instrumentation schematic of Pasteurizer II under minimally and maximally adequate instrumentation. Principal pipe legs are labeled, and temperatures, volume flow rates, vat volumes, and setting points are noted at the appropriate places.

Pasteurizer II heats milk ("feedstock") from an input temperature of 4.44°C to a temperature in the range of 76.7°C to 82.2°C. Milk travels through the system at a minimum fluid flow rate of 725 L/hr and must be held in the target temperature range for 15 s to be legally pasteurized (Hall & Trout, 1968). A holding tube helps to meet this requirement.

Milk enters Pasteurizer II through the input leg (pipe) into the input vat. If the input vat becomes too full, excess milk is removed into the overflow vat. From the input vat, the milk travels through the first feedstock pipe (Feedstock Leg A) to the feedstock pump, which pumps the milk through Feedstock Leg B to the regenerator heat exchanger, which "preheats" milk when the system is producing pasteurized milk and sending it to the product vat. The preheated milk then flows through Feedstock Leg C to the high-temperature short-time heat exchanger (HTST), where it is further heated to the pasteurization temperature range by hot water heated by the water heater and circulated through the HTST by the water pump. The feedstock goes through the holding tube so that it is held at pasteurization temperature for the required time and is then sent to the three-way valve via the last feedstock pipe (Feedstock Leg D).

At the three-way valve, the feedstock is diverted according to its temperature. If the temperature is between 76.7°C and 82.2°C, feedstock travels through the regenerator to the product vat (the "producing" state); if the temperature is below 76.7°C, then it is recycled through the recycle leg to the input vat (the "recycling" state); and if the temperature is above 82.2°C, then it is sent out of the system through the burnt feedstock pipe to the burnt vat (the "burning" state). Pasteurizer II can be difficult to control if it transitions from the producing state into the recycling or burning state because of the danger of overcorrections by the human operator. Overcorrections are exacerbated by heating lags and the delay in the holding tube. A naive operator will produce oscillations in and out of these states.

Pasteurizer II simulates "normal" sensor noise on each displayed variable. The value displayed and used for calculating any higher-order variables is a normal random variate (Law & Kelton, 1991) between ±5 standard deviations of the value of the variable from the simulation mean itself. Standard industrial component tolerances from Internet component distributors were used to set the value at ±5 standard deviations (see Reising, 1999, for details). Therefore, during steady-state conditions, all digital sensor readouts and graphical elements vary slightly across the simulation's 2-s updates.

Interfaces for Pasteurizer II

EID interface. The development of the ecological interface (EID interface) for Pasteurizer II is described fully in Reising (1999) and in Reising and Sanderson (2002a). Here we provide a brief description of the EID interface and contrast it with the piping-and-instrumentation diagram interface (PID interface) used as a control condition. The specific EID and PID interfaces developed for this study were each just one sample of their category, and other variants are possible.

The design of the EID interface followed the principles of EID by representing information about Pasteurizer II from all levels of the abstraction hierarchy and by supporting the operator at the least loading level of cognitive control. Figure 3 is a sensor AH of Pasteurizer II that shows four levels of abstraction: functional purpose, priorities/values, purpose-related functions, and object-related functions. Because Pasteurizer II is a simulation, the fifth level, physical objects, is not shown. The names of the five displays in the ecological interface are

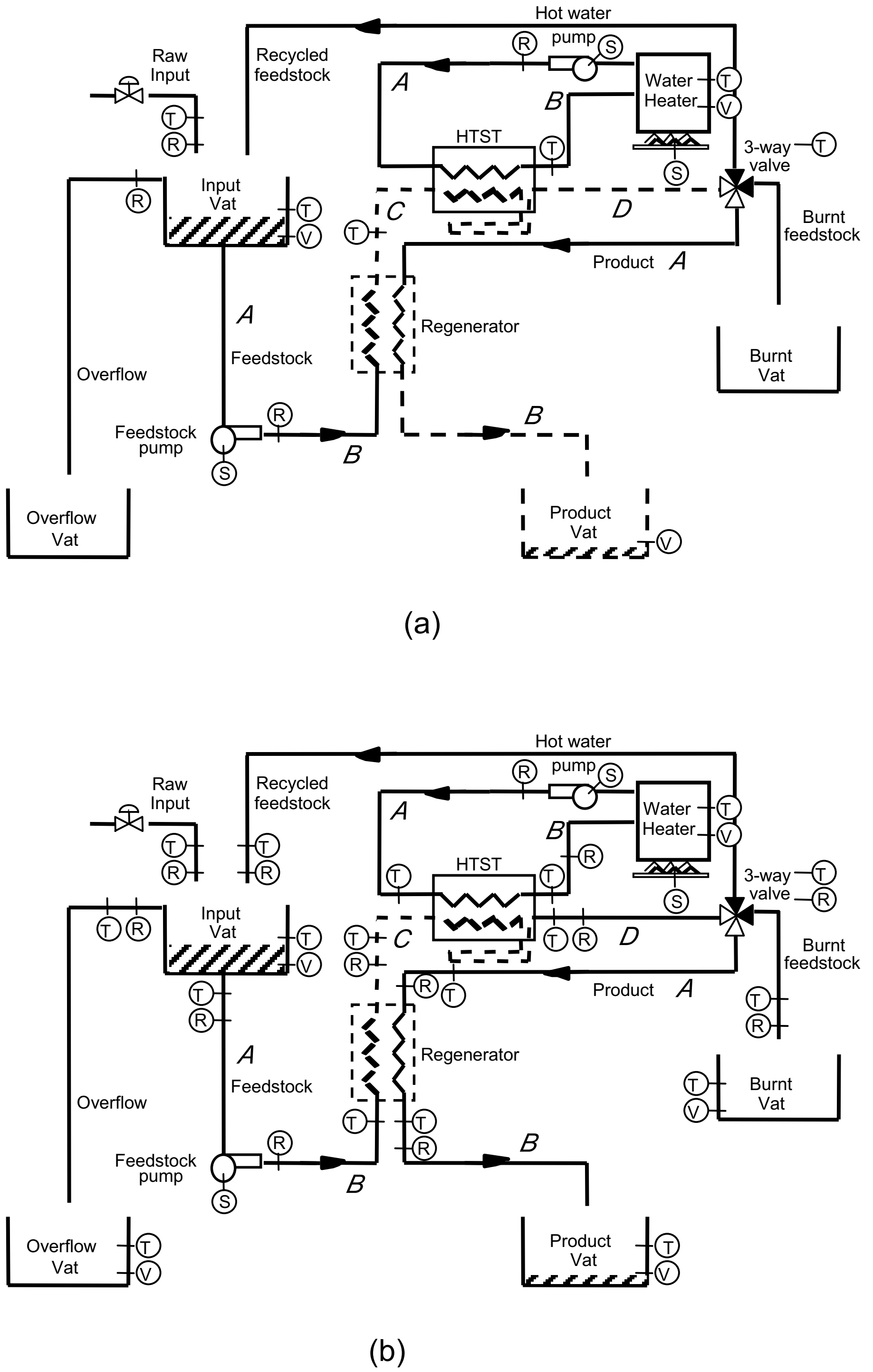

Figure 2. Configuration of Pasteurizer II microworld. (a) The minimally adequate sensor set. (b) The maximally adequate sensor set. Temperature (T), volume (V), and flow rate (R) values are identified by the general area of Pasteurizer II in which they reside (e.g., feedstock, HTST), type of component (vat, pipe), and identifier for the relevant part of the component (A, B, C, D). S = control setting. Dotted lines illustrate the impact of a failure of the feedstock temperature sensor at Leg C.

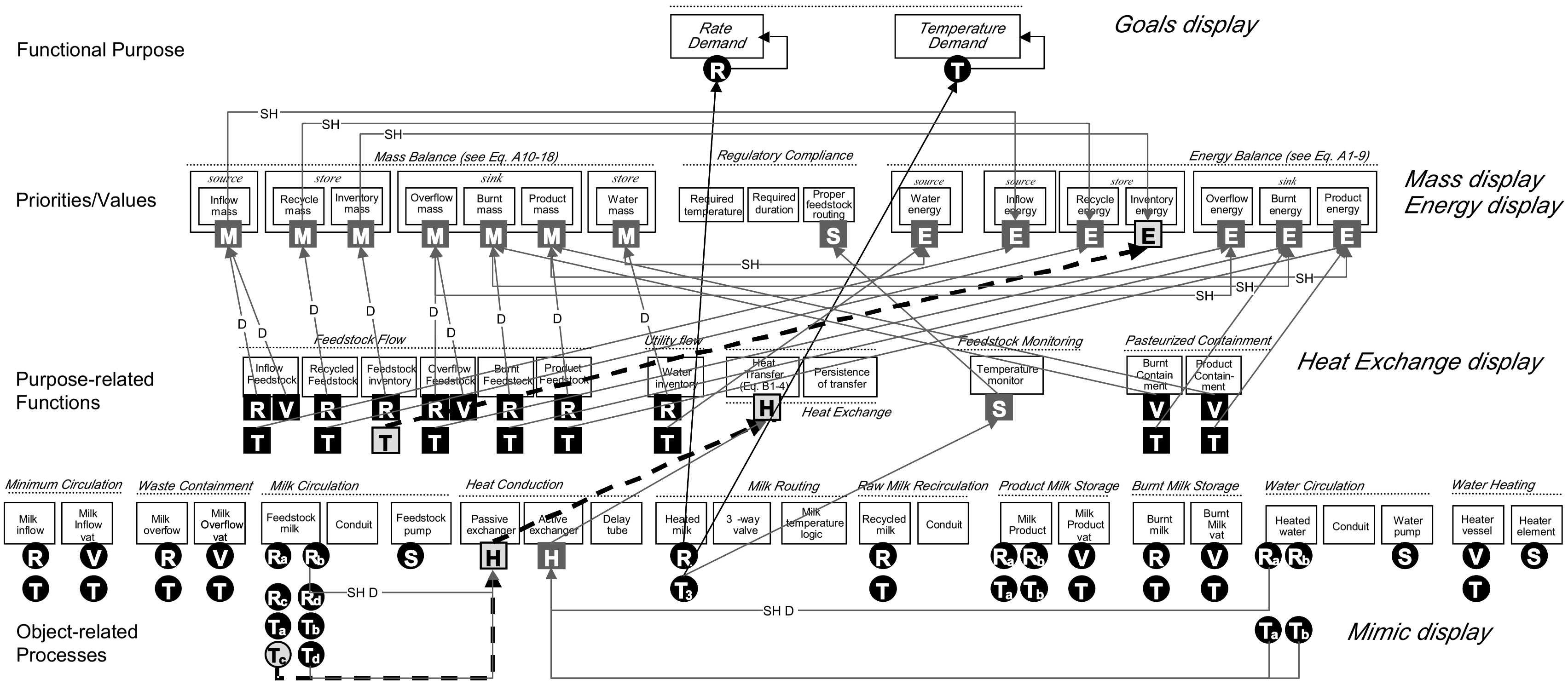

Figure 3. Sensor abstraction hierarchy of Pasteurizer II with the maximally adequate instrumentation configuration. Black circles are physical sensors (or control settings if S), black squares are values directly read from a sensor, and gray squares are calculated values. Dashed lines show propagating derivational effects of the feedstock temperature sensor failure. (Adapted from *International Journal of Human-Computer Studies,* Vol. 56, D. C. Reising and P. M. Sanderson, "Work Domain Analysis and Sensors: II. Pasteurizer II Case Study," p. 626. Copyright 2002, with permission from Elsevier.)

placed next to the level of abstraction from which they are drawn.

The ecological interface is shown in Figure 4. At top left, functional purpose is represented in the goals display (the bar graphs); here the temperature and volume flow rates being demanded of the pasteurizer are shown.

Below the goals display, priorities/values are shown – in this case, as mass and energy flow rate displays. The display at center left is the mass display and the tall display to its right is the energy display. The energy flow rate display depicts the energy flow at each leg of the feedstock and water loops, preserving something of the topographic structure of Pasteurizer II. The parallelism of diagonal lines representing the relationship between hot and cold leg energies of the heat exchangers gives the viewer a quick indication of whether the energy balance is preserved. The mass flow rate display depicts mass flow for each leg of Pasteurizer II's feedstock and water loops using a topographic structure consistent with that of the energy display. When all bars within each loop are at the same height, the mass balance is preserved.

Along the bottom of the EID interface, purpose-related functions are captured in the heat exchange display. The four panels of the heat exchange display relate the flow rate of fluid going through each side of the heat exchangers to the temperatures of the fluid and the efficiency of the heat exchanges themselves. The two left panels show these relationships for the regenerator, and the two right panels show these relationships for the HTST exchanger. Finally, object-related processes are represented at upper right in a mimic display, in which the physical layout of Pasteurizer II is reproduced.

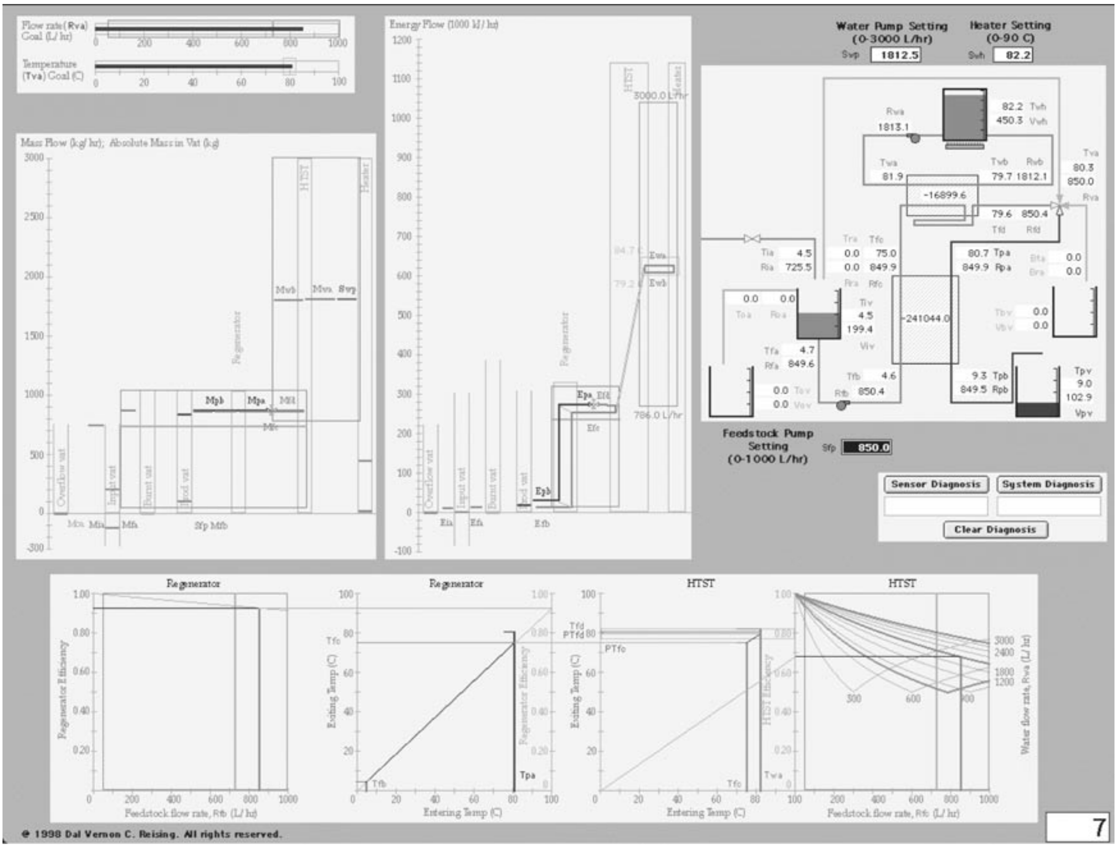

Figure 4. EID (ecological) interface with maximally adequate sensor set. (Image was displayed on a 19-inch [48-cm] monitor; this figure is provided for illustrative purposes only, and details are not meant to be legible.)

Subsystems and key relations within Pasteurizer II, together with their labels, are color coded within each interface. Target values and boundaries are also represented and color coded. First, a bright green line in both the goals and heat exchange displays indicates the target feedstock flow rate minimum of 725 L/hr. The energy flow rate and mass flow rate that are equivalent to 725 L/hr, taking into account current conditions, are shown in bright green in the mass and energy flow rate displays. Second, the upper and lower boundaries of a yellow rectangle indicate the maximum and minimum energy flows, given the heater temperature setting, that keep the temperature at the three-way valve within the goal region for the system. This temperature goal region is also noted in yellow on the goals display.

The arrangement of displays in the EID interface reflects the general spatial layout of the abstraction-decomposition hierarchy, as recommended by Dinadis and Vicente (1996). All levels of abstraction are simultaneously present but are spatially separated on the screen. In other words, the interface has high temporal but low spatial integration (Burns, 2000). We did not attempt further spatial integration because our initial focus was on designing and evaluating effective representations for the different levels of abstraction, exploring the layouts proposed by Dinadis and Vicente (1996) and Jamieson (1998).

PID interface. The PID interface was a larger version of the mimic display within the EID interface (see Figure 5). Although the mimic display shows object-related processes, it also represents some properties of Pasteurizer II at the purpose-related functions levels of the abstraction hierarchy, such as feedstock and water flow rates.

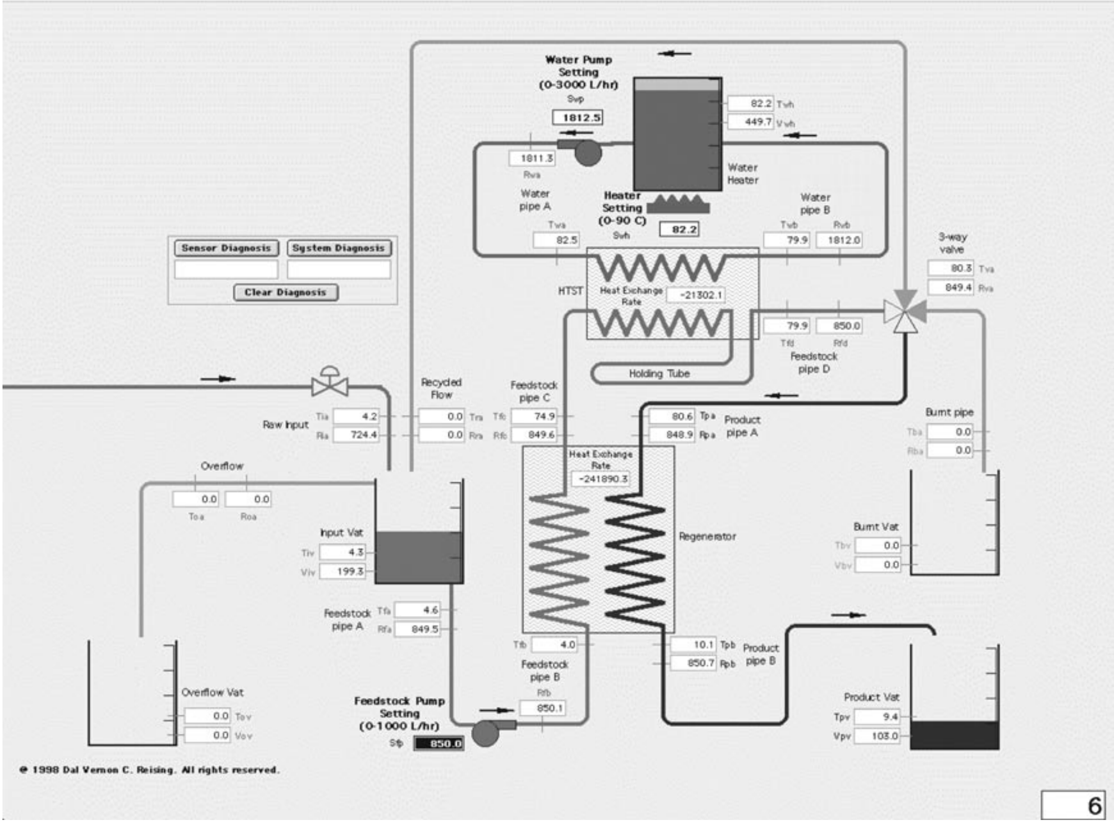

Figure 5. PID (mimic) display with maximally adequate sensor set. (Image was displayed on a 19-inch [48-cm] monitor; this figure is provided here for illustrative purposes only, and details are not meant to be legible.)

Information at the purpose-related functions level of abstraction has not usually been included in the PID interfaces used in studies of ecological interface design. However, several researchers (e.g., Maddox, 1996) have criticized this, claiming that in practice it would be highly unusual for a PID interface not to include flow information and that comparing a PID interface without flow information with an EID interface that does include flow information is unfair. Although this point can be debated, we sought to deflect this potential criticism by including flow information in our PID interface (see Figure 5). This would make it more difficult to find a significant difference between our PID and EID interfaces, but any superiority of our EID interface would have been gained in a very conservative test. In actual practice, configural displays within an ecological interface may be based on information at a given level of abstraction but may be supplemented with information from other levels for ease of reference (see Bennett & Walters, 2001). In this manner we set up a relatively conservative comparison of EID and PID interfaces.

Maximum and Minimum Sensor Configuration Manipulation

The number of sensors providing information about Pasteurizer II's state was varied to provide minimally and maximally adequate information, and the result is shown in Figure 2. The maximally adequate set was determined by finding the set needed to derive each mass and energy flow rate value for each component in Pasteurizer II independently of all other components. Figure 1 illustrates the terminology and logic. Specifically, the maximally adequate set of sensors allows a process variable *Y* (e.g., mass flow rate) to be calculated for each individual piece of equipment (e.g., each pipe leg) from an individual measurement of process variable *X* (e.g., volume flow rate) on each piece of equipment. Note that the kind of redundancy that results is different from the kind of redundancy normally discussed in sensor verification and validation, which is concerned with confirming the value of *X* by having multiple, independent measures of *X* at the same location in the process rather than at successive locations, as in our case. The maximally adequate set of sensors was considered to be derivationally adequate for displaying higher-order information independently for the various pieces of equipment.

To determine the minimally adequate set, we removed sensors from the maximal set if, under normal steady-state conditions, the equivalent upstream or downstream process variable would be equal to the measured process variable in question. Sensors were removed until a minimal set remained from which all information in the abstraction hierarchy could still be derived.

Once the sensors were placed on the Pasteurizer II sensor AH, whether in the minimally or maximally adequate sensor configuration, the sensor AH indicated the information to be collected and the derivations to be performed in order to build the ecological interface. The sensor AH also showed how the effects of failures would propagate. For example, Figure 2b shows the impact of a failure of the temperature sensor at Feedstock Leg C under maximally adequate instrumentation, and Figure 3 is the associated sensor AH that shows the effect on higher-level properties. An equivalent sensor AH developed for the minimally adequate sensor set in Figure 2a let us predict when participants might experience difficulties in performing failure diagnosis. (See Reising & Sanderson, 2002c, for a full analysis.)

As Figures 2a and 2b show, the impact of the failure is relatively localized under maximally adequate instrumentation but more widespread under minimally adequate instrumentation. This contrast will be even more evident in a sensor AH for minimally adequate instrumentation (see Reising and Sanderson, 2002c, for an example) and is evident with derived values in a configural display, making it harder for the viewer to pinpoint the source of the problem under minimally adequate instrumentation. For example, the energy display in Figure 6b shows that under maximally adequate instrumentation, only the display element related to Feedstock Leg C moves out of line (see dark arrow). In contrast, Figure 6a shows that under minimally adequate instrumentation, not only the display element related to Feedstock Leg C but also the element related to Product Leg B are affected, producing a motion in the display that is falsely consistent with a normal state change. A more detailed

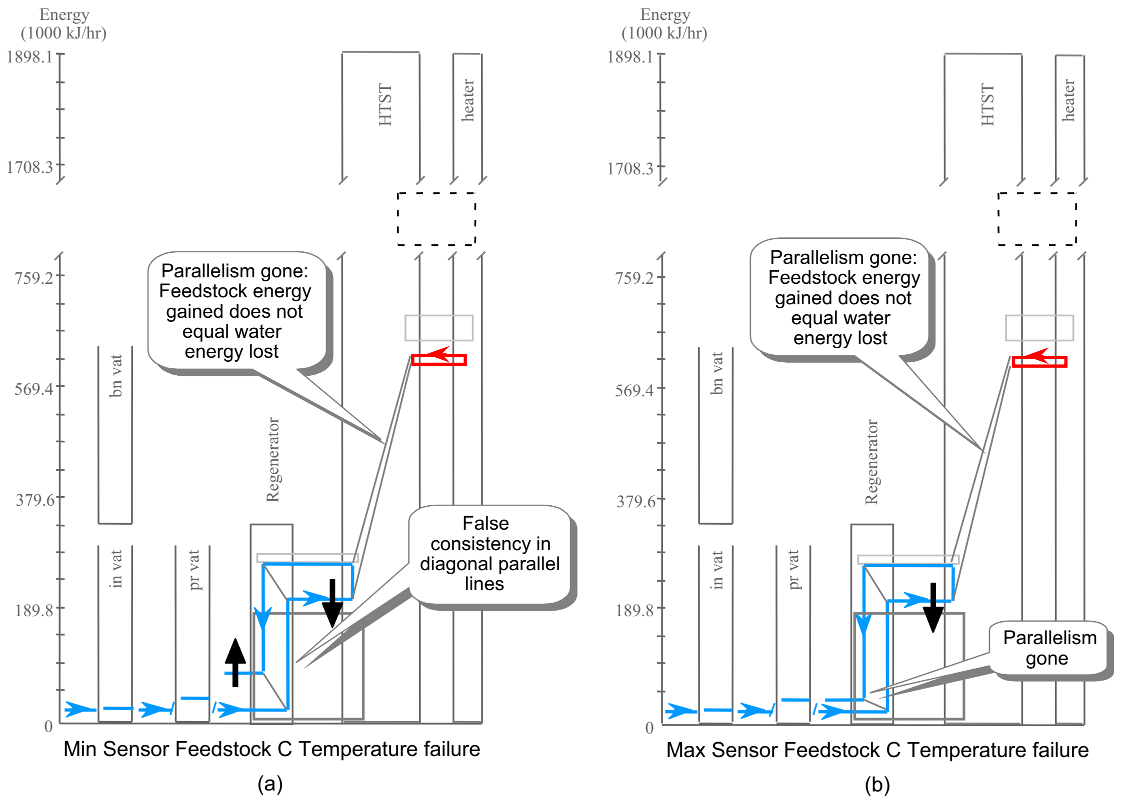

Figure 6. Schematic of the energy display showing effects of the feedstock temperature sensor failure. (a) Minimal instrumentation. (b) Maximal instrumentation. Dark arrows show direction of change, and callouts highlight features (neither was shown on the actual display). (Adapted from *International Journal of Human-Computer Studies,* Vol. 56, D. C. Reising and P. M. Sanderson, "Work Domain Analysis and Sensors: II. Pasteurizer II Case Study," p. 627. Copyright 2002, with permission from Elsevier.)

description with further examples is given in Reising and Sanderson (2002c).

Experimental Hypotheses

In this experiment we tested the following hypotheses. H1: Failure diagnosis performance should be better under maximally adequate instrumentation than under minimally adequate instrumentation. H2: Failure diagnosis performance should be better with the EID interface than with the PID interface. H3: When instrumentation goes from being maximally adequate to being minimally adequate, failure diagnosis performance with an EID interface will suffer more than failure diagnosis performance with a PID interface will (interface and instrumentation will interact).

METHOD

Participants

The 44 participants were senior undergraduate or graduate students in the Departments of Mechanical and Industrial Engineering, Food Science and Human Nutrition, and Nuclear Engineering at University of Illinois at Urbana-Champaign. Participants were screened for red-green color blindness. Each participant had completed at least one university-level engineering subject in thermodynamics. Participation was voluntary, and each participant was paid $5.00/hr for his or her time. Participants were randomly assigned to one of four combinations of interface and instrumentation in such a way that level of education, gender, and major field of study were controlled across conditions.

Apparatus and Stimuli

Pasteurizer II was implemented in C using Metrowerks® CodeWarrior® IDE and ANSI C programming language libraries. The interfaces were developed using Metrowerks® PowerPlant™ C++ object class libraries. Pasteurizer II was run on an Apple® Power Macintosh® G3 233 MHz machine with 32 MB of RAM and a disk cache set at 1024 K. MacOS Version 8.1 was the operating system running on the G3. The computer monitor was a 19-inch (48-cm) Apple® color monitor (18 inches, or 46 cm, viewable) with the color depth set to 256 colors, the resolution set to 1152 × 870 pixels at 75 Hz, and the gamma set to Mac Std. A standard Apple® keyboard and mouse were used for data entry.

Design

A between-within design was used, with the independent variables of interface (EID vs. PID) and instrumentation (minimum vs. maximum) manipulated between participants. This resulted in four conditions: EID.max, EID.min, PID.max, and PID.min. There were 11 participants in each condition. Controlled variables manipulated within participants were the kind of failure (sensor vs. system) and failure diagnosis sessions, or days (Day 4 and Day 5). Across Days 4 and 5, participants experienced 16 failures altogether: 6 system failures and 10 sensor failures.

The EID.max interface is shown in Figure 4, and the PID.max interface is shown in Figure 5. The EID.min and PID.min interfaces were exactly the same as their maximum counterparts, except that fewer data points were shown on the mimic diagram in each case because there were fewer sensors in the minimum configuration.

Procedure

The experiment involved five 1-hr experimental sessions that were usually run on separate days (the very few exceptions meant that the experiment was completed in 4 days with a minimum 4-hr separation between same-day sessions. However, the fourth and fifth sessions were never on the same day). Days 1, 2, and 3 involved training, during which participants were asked to achieve operating goals in exercises of gradually increasing complexity, using the interface they would also use on Days 4 and 5. On Days 4 and 5, participants were required to diagnose failures that occurred during each of the eight trials each day. Participants were asked to “maintain production if possible while making a diagnosis quickly and accurately.” Using the diagnoses box on the Pasteurizer II screen, participants entered their hypotheses by clicking the sensor or system failure button, clicking on the screen element that had failed, and then reviewing the resulting diagnosis in a text field. Participants could change the diagnosis if they wished. Trials always lasted 7 min, regardless of when a diagnosis was made or its accuracy. See Reising (1999) for further details of the experimental procedure.

At the end of the five sessions with Pasteurizer II, participants in the EID conditions answered a questionnaire in which they rated each of the five configural displays and the elements of each display (its boundaries, symmetries, relationships, etc.) in terms of their usefulness for Pasteurizer II control and failure diagnosis. All participants were also asked to give open-ended comments about the most important sources of information they used for controlling the system.

RESULTS

First, the performance of participants at failure diagnosis with the different displays will be presented as well as further analyses that help to characterize their failure diagnosis performance. Second, participants’ responses to questionnaires will be presented.

Failure Diagnosis Performance

Correct diagnoses. Correct diagnoses were cases in which failures were correctly classified as sensor or system failures and correctly attributed to a particular component. The percentage of correct failure diagnoses made on Days 4 and 5 was analyzed in a between-within analysis of variance (ANOVA) with the between-subjects factors of interface (EID vs. PID) and instrumentation (maximum vs. minimum) and the within-subjects factors of failure (sensor vs. system) and days (Day 4 vs. Day 5).

Hypothesis 1 was that there should be better failure diagnosis performance in the maximum condition than in the minimum condition. This was supported by a main effect of instrumentation, *F*(1, 40) = 11.44, *p* < .01, which showed that failure diagnosis was performed more accurately with maximally than with minimally adequate instrumentation.

Hypothesis 2 was that performance with the EID interface would be better than with the PID interface. However, there was no main effect of interface, *F*(1, 40) = 0.09, *ns,* largely because of the Interface × Instrumentation interaction. Therefore we made the same test under maximal instrumentation conditions. In a directional test of EID.max versus PID.max within the Interface × Instrumentation interaction using a Newman-Keuls test, EID.max was found to produce significantly better failure diagnosis performance, *p* < .05, one-tailed. Although the direction of this effect was reversed under minimal instrumentation conditions, a Newman-Keuls test showed that performance with EID.min was not significantly worse than with PID.min.

Hypothesis 3 was that performance with the EID interface would show a greater drop from the maximum to the minimum instrumentation condition than would performance with the PID interface. This was supported by a significant interaction of Interface × Instrumentation, *F*(1, 40) = 6.13, *p* < .05. The two EID conditions were at the extremes (see Figure 7). A directional Newman-Keuls test within the Instrumentation × Interface interaction indicated that EID.max supported much better failure diagnosis performance than did EID.min, *p* < .01, one-tailed.

A main effect of day indicated an overall improvement in failure diagnosis performance from Day 4 to Day 5, *F*(1, 40) = 11.5, *p* < .01. Moreover, a main effect of failure indicated that sensor failures were diagnosed more accurately overall than were system failures, *F*(1, 40) = 25.9, *p* < .001. Participants using the minimum interface found some of the system failures on Day 5 particularly hard to diagnose, which led to some interactions with day and instrumentation. For example, the Day × Interface × Instrumentation interaction reached significance, *F*(1, 40) = 7.1, *p* < .05, as did the Day × Failure × Instrumentation interaction, *F*(1, 40) = 18.6, *p* < .001. Each of the three system failures on Day 5 involved either a pump failure or a leak in a pipe, all of which required a flow sensor on

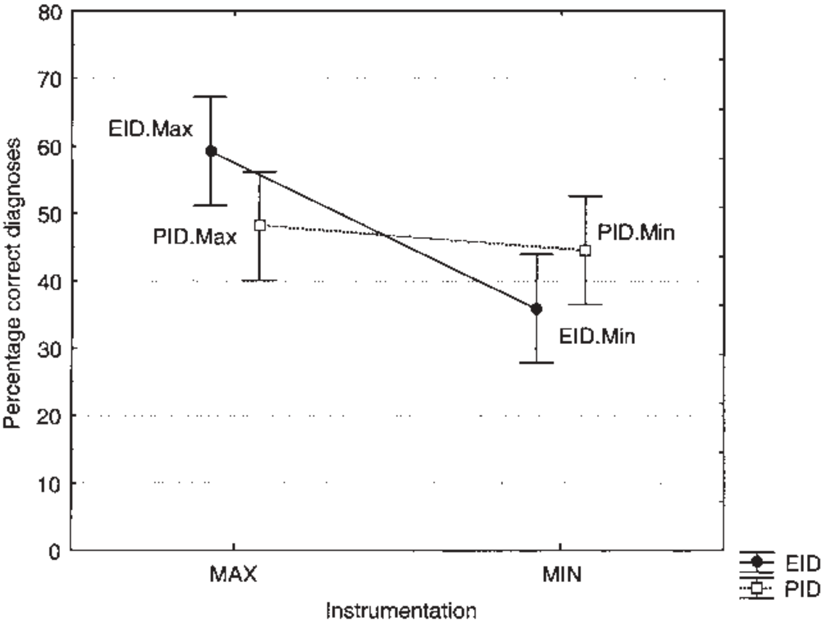

Figure 7. Percentage correct failures diagnoses for interface and instrumentation. Error bars are 95% confidence intervals.

a subsequent downstream pipe segment to diagnose without highly sophisticated reasoning; however, that sensor was available only in the maximal instrumentation conditions. The system failures on Day 4 involved leaks in the vats and water heater, and therefore all four conditions had the same primary instrumentation needed to diagnose the failures.

Time to diagnose failures. Our primary goal was to analyze failure diagnosis accuracy, but we also analyzed diagnosis time. There were three phases of failure diagnosis: time after the onset of the failure to detect anomalous behavior (Phase 1), time after detection to classify the anomalous behavior as a system or sensor failure by clicking the sensor or system failure button on the screen (Phase 2), and time after classification to identify the actual failed component by clicking the failed component on the mimic display and to confirm verbally that it was the final diagnosis (Phase 3). Unfortunately, we did not collect events that reliably marked the end of Phase 1.

When we examined the time it took participants either to classify the kind of failure (Phases 1 and 2) or to confirm their final diagnosis (Phases 1, 2, and 3), there were no significant effects for interface, instrumentation, or day. However, when we examined times on trials that led to correct responses only, for Phases 1 and 2 we found a selective speed-up of responding for the EID.max condition alone from Day 4, when it was the slowest of all conditions observed (57.5 s), to Day 5, when it was the fastest of all conditions (33.0 s), the Interface × Instrumentation × Days interaction reaching significance, $F(1,40) = 6.12$, $p < .05$. Moreover, the time to complete Phase 3 correctly was negatively correlated with time to complete Phases 1 and 2: For Day 4, $r(42) = -0.31$, and for Day 5, $r(42) = -0.27$, $.10 > p > .05$. This indicated that the more quickly participants made correct classification judgments (Phases 1 and 2), the more slowly they confirmed their final diagnosis (Phase 3), and vice versa. An Interface × Instrumentation interaction for Phase 3 supported this interpretation, $F(1, 40) = 6.9$, $p < .05$. Overall, however, there are no strong effects of instrumentation or interface on participants' time to respond to failures. These factors appear to affect accuracy instead.

Questionnaire Results

The questionnaire answers revealed participants' subjective impressions of the usefulness of the displays within the EID interface under the minimal and maximal instrumentation conditions and their impressions of how important each display was for control.

Usefulness. Participants' ratings on a scale from 1 to 10 for the usefulness of the five displays making up the EID interface are given in Figure 8. A between-within ANOVA was performed on the ratings using the between-subjects factor of instrumentation (minimum vs. maximum) and the within-subjects factor of display (goal, mass, energy, heat exchange, and mimic). Perceived usefulness ratings were much higher for the EID.max than for the EID.min interface, $F(1, 20) = 8.79$, $p < 0.01$, and perceived usefulness varied markedly across displays, $F(4, 80) = 20.4$, $p < .001$. In addition, there was a marginal interaction of Instrumentation × Display, $F(4, 80) = 2.2$, $p = .077$. The mimic and goals displays were rated as equally highly useful by participants in EID.max and EID.min conditions, but participants in the EID.min condition tended to rate the mass, energy, and heat exchange displays as less useful than participants in the EID.max condition did.

Importance for control. Participants were also asked to comment on the interface features that were most useful for controlling Pasteurizer II. In the EID conditions, most participants mentioned two different features, and the collated results are shown in Table 1. Both the EID.max and EID.min participants considered the mimic display the most important display for control – not surprising, given that control and diagnosis actions were made on the mimic display in all conditions. However, for the EID.max participants the energy display was perceived to be the second and the mass display the third most highly valued display for control, whereas for EID.min participants the goals and energy displays were perceived to be equally important after the mimic display. The goals display simply showed the outcome of participants' control and was not integrated with any controllable variables. Chi-square statistics could not be performed on these data because of the nonindependence of observations across

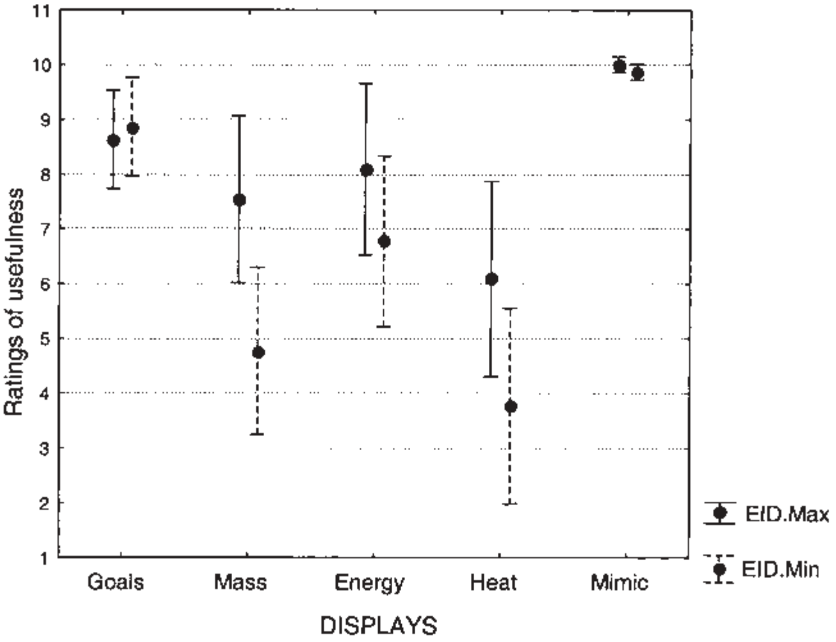

Figure 8. Overall ratings for usefulness of displays within the ecological interface, across minimally adequate and maximally adequate instrumentations of the interface. Error bars are 95% confidence intervals.

cells, but the results suggest that EID.max participants made greater use (14 mentions) of the configural displays included in the ecological interface than did EID.min participants (7 mentions).

For participants using the PID interfaces, comments about features they found useful were limited to properties of the mimic display because it was the only display they saw. Participants considered certain state variables important for control, including (almost universally) the temperature at the three-way valve and the settings of the pumps and the heater. Somewhat poignantly, in the PID.min condition, in which the display is least helpful, one participant commented, “a calculator would be nice” – precisely the kind of symbolic, high-workload tool for which EID should remove any need.

DISCUSSION

In response to a long-standing need in the area of EID, this experiment addressed the issue of whether inadequate instrumentation compromises the effectiveness of interfaces designed according to EID principles (Reising & Sanderson, 2002b, 2002c; Vicente, 2002; Vicente & Rasmussen, 1992). Participants with theoretical domain knowledge worked with versions of Pasteurizer II equipped with instrumentation

TABLE 1: Number of Times a Display Was Noted as Important for Control of Pasteurizer II

		Display Within the EID Interface				
	n	Goals	Mass	Energy	Heat	Mimic
EID.max	11	2	5	8	1	9
EID.min	11	5	2	5	0	9

Note. Data taken across all participants in each condition. Each display is counted once for each participant, but each participant mentioned approximately two displays.

that was either maximally or minimally adequate for representing information needed at all levels of the AH and in which information was displayed in either an EID or a PID interface. With maximally adequate instrumentation the EID interface supported better failure diagnosis performance than did the PID interface, whereas with minimal instrumentation the benefits of the EID interface were lost entirely and failure diagnosis performance with the EID interface became nonsignificantly worse than with the PID interface. We found not only the hypothesized interaction between instrumentation and interface but also that the EID interface based on maximally adequate instrumentation supported the most accurate failure diagnosis performance of all four conditions, whereas the EID interface based on minimally adequate instrumentation supported the least accurate performance.

In summary, the advantages of an EID interface appear to be confined to situations in which there is maximally adequate instrumentation. Where there is only minimally adequate instrumentation, neither the PID nor the EID interface is reliably better than the other.

Interface and instrumentation appeared to affect diagnosis accuracy more than they affected diagnosis speed. There were no strong effects of interface or instrumentation on any phase of failure diagnosis that we measured directly. However, participants using the ecological display with maximally adequate instrumentation became quicker over days at correctly classifying failures as sensor or system failures. These results suggest that although participants were benefiting from high-level information in the interface when performing diagnoses, they were still learning how to extract that information efficiently.

Questionnaire results indicate that participants using the EID interface with maximally adequate instrumentation considered the mass, energy, and heat transfer displays more useful and more important for control than did participants using the EID interface with minimally adequate instrumentation. The mass, energy, and heat transfer displays were the ones that relied on sensors to derive higher-order information. In contrast, the perceived usefulness and importance for control of the mimic and goal displays was the same for EID interface participants, regardless of whether their instrumentation was minimally or maximally adequate.

Our results give a conservative indication of display usefulness and importance for control. First, all participants used the mimic display to make control inputs and to note their final diagnoses, regardless of where changes or failures were first noticed. Despite this, EID participants experiencing maximally adequate instrumentation reported that the displays presenting higher-order information were more useful than did EID participants experiencing minimally adequate instrumentation. Second, state changes became evident in the digital readouts of the mimic display before they became evident graphically in the energy, mass, and heat transfer displays because of the fidelity of the digital readouts versus the graphics (one pixel in the mass display was equivalent to 8.33 kg/hr, whereas volume flow rate readouts in the mimic display were to the 10th decimal place). Digital readouts supplementing rather than replacing higher-order information can further enhance performance with EID displays (Bennett & Walters, 2001), but we did not use them.

Our results indicate that the EID interface with maximally adequate instrumentation supported better performance than did the equivalent PID interface, despite a somewhat conservative test. As noted earlier, we included flow information in our PID interface. In most other contrasts of EID with PID interfaces (e.g., the dual reservoir system [DURESS] physical vs. physical-plus-functional interface comparisons; Vicente, 2002), flow information was not included in the PID interface. However, some researchers have questioned the validity of such results (Maddox, 1996). We included flow rate in our PID interface even though it was not strictly necessary by some interpretations and even though it reduced the possibility of finding a difference. Moreover, in light of the criticism that interface contrasts should be based on the same underlying sensor set, we also standardized the instrumentation used when we compared EID and PID interfaces: Both were either maximally adequate or minimally adequate. These manipulations may have attenuated the distinction between EID and PID interfaces, especially as Ham and Yoon (2001a, 2001b) have demonstrated that the advantages of EID interfaces can be demonstrated incrementally as successive levels of even nongraphical higher-order information are added to an interface.

The interaction between instrumentation and interface eliminated any overall superiority for the EID interface. In the past, the failure to achieve strong EID interface superiority effects has been attributed to failing to test the interface with unanticipated failures (Ham & Yoon, 2001a, 2001b). Our results suggest that researchers also need to be careful that EID interface effects have not been attenuated by inadequate instrumentation for the information derived.

Participants apparently did not consider all higher-order displays of the present EID interface to be as effective in providing useful information as the mimic display, although participants may have been biased toward the mimic display because it was also used in all conditions for control and entering diagnoses. Briefly, participants considered display features valuable if they gave simple visual information about higher-level properties and boundaries, as has been found elsewhere (Ham & Yoon, 2001a, 2001b; Janzen & Vicente, 1998; Pawlak & Vicente, 1996; Sakuma, Itoh, Yoshikawa, & Monta, 1995). The heat exchange display was perceived as least useful either because there were no instructions for interpreting the display, because the display was not spatially integrated with others, or because participants could use information in other displays to solve most failure diagnosis problems. Overall, the ecological interface combined more- and less-successful properties (see Reising & Sanderson, 2002a, for further details).

In summary, although our conservative approach may have attenuated the contrast we would have otherwise expected to see between performance with the EID and PID interfaces with maximally adequate instrumentation, our results still bear out the superiority of the EID interface over the PID interface for failure diagnosis under these conditions.

Limitations of the Present Study

The present study has several limitations that point to possible follow-up studies. First, training was relatively short. Participants may have not reached asymptotic control performance and may not have had a chance to understand the configural displays as thoroughly as they might have.

Second, we chose failures that did not initially drive Pasteurizer II out of the producing state, so that participants would not become locked up on control rather than diagnosis. Failures with more extreme consequences, however, would allow us to study the effect of instrumentation on control performance.

Third, a pasteurization plant is a relatively simple industrial system as compared with systems such as catalytic cracker units in a refinery or the steam generation side of pressurized water reactor nuclear stations. However, the Pasteurizer II simulation is based closely on a real pasteurization system, which it simulates with considerably more fidelity than is typical for most laboratory microworlds. Therefore, Pasteurizer II is sufficiently complex for a first investigation of the questions asked, and we look forward to the scalability issue being investigated in future studies. Other researchers are applying EID to industrial-scale applications that range from 250 to 400 process variables (Burns, 2000; Jamieson, 2002, 2003), including nuclear applications (Ham & Yoon, 2001a; Yamaguchi & Tanabe, 2000). In another departure from realism, our participants were not subjected to the additional distractions and responsibilities beyond process monitoring that might be required in an industrial setting, and they did not have the backup of auditory alarms.

Fourth, we have tested our hypotheses with just one ecological interface and one general class of instrumentation inadequacy. Future work should investigate the impact of minimally adequate instrumentation when an ecological interface consists of a different arrangement of display elements. For example, Ham & Yoon (2001b) have shown benefits when display elements as simple as bar graphs and text fields are arranged according to a mean-ends structure. Moreover, following Burns (2000), we could seek to achieve spatial as well as temporal integration of our five displays. Such investigations would let us tease apart the relative impact of information content and information form on the effects seen in the present experiment. Finally, other forms of instrumentation inadequacy should be explored.

Implications for Interface Design

Overall, our results suggest that if ecological interfaces are to be interpretable and effective when needed, then they should be free of ambiguities that arise when their sensors are only minimally adequate to derive and display higher-order system properties. Tools such as functional information profiles (Burns & Vicente, 1996), variable measurement inventories (Jamieson, 1998), and sensor abstraction hierarchies (Reising & Sanderson, 1996, 2002b, 2002c) are available to help the analyst ensure that all information that should be displayed is displayed and that all displayed information is free of underlying dependencies that might leave it prone to misinterpretation. In addition, display programming logic may need to become more intelligent about the interaction between sensor information and forms of visual representation (Leveson, 1995). In summary, our results highlight the need to develop EID interface engineering techniques to provide interface verification techniques alongside EID interface design principles.

ACKNOWLEDGMENTS

We would like to thank Amanda Stewart for her assistance in conducting the experimental evaluations. In addition, we thank Chris Wickens for his contributions to the experimental design as well as to the control of features between the EID and PID interfaces. We also gratefully acknowledge the Department of Mechanical and Industrial Engineering, the Institute of Aviation, and the Research Board at the University of Illinois at Urbana-Champaign for providing financial support to these studies.

REFERENCES

Beltracchi, L. (1998). *Encoding a model-based display with dynamic data.* Unpublished manuscript.

Bennett, K. B., Nagy, A. L., & Flach, J. (1997). Visual displays. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (2nd ed., pp. 659–696). New York: Wiley.

Bennett, K., & Walters, B. (2001). Configural display design techniques considered at multiple levels of evaluation. *Human Factors, 43,* 415–434.

Bisantz, A., Roth, E., Brickman, B., Lin, L., Hettinger, L., & McKinney, J. (2003). Integrating cognitive analyses in a large-scale system design process. *International Journal of Human-Computer Studies, 58,* 177–206.

Burns, C. M. (2000). Putting it all together: Improving integration in ecological displays. *Human Factors, 42,* 226–241.

Burns, C. M., & Vicente, K. J. (1996). Comparing the functional information content of displays. In *Proceedings of the 28th Annual Conference of the Human Factors Association of Canada* (pp. 59–64). Mississauga, Canada: Human Factors Association of Canada.

Dinadis, N., & Vicente, K. J. (1996). Ecological interface design for a power plant feedwater system. *IEEE Transactions on Nuclear Science, 43,* 266–277.

Hall, C. W., & Trout, G. M. (1968). *Milk pasteurization.* Westport, CT: AVI.

Ham, D.-H. & Yoon, W. C. (2001a). Design of information content and layout for process control based on goal-means domain analysis. *Cognition, Technology, and Work, 3,* 205–223.

Ham, D.-H., & Yoon, W. C. (2001b). The effects of presenting functionally abstracted information in fault diagnosis tasks. *Reliability Engineering and System Safety, 73,* 103–119.

Hayter, D. (1996). *An application of ecological interface design: Boiler level control.* Unpublished manuscript, University of Toronto, Canada.

Jamieson, G. A. (1998). *Ecological interface design for petrochemical processing application.* Unpublished master's thesis, University of Toronto, Canada.

Jamieson, G. A. (2002). *Ecological interface design for petrochemical process control: Integrating task- and system-based approaches.* Unpublished Ph.D. thesis, University of Toronto, Canada.

Jamieson, G. A. (2003). Bridging the gap between cognitive work analysis and ecological interface design. In *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting* (pp. 273–277). Santa Monica, CA: Human Factors and Ergonomics Society.

Jamieson, G. A., & Vicente, K. J. (2001). Ecological interface design for petrochemical applications: Supporting operator adaptation, continuous learning, and distributed, collaborative work. *Computers and Chemical Engineering, 25,* 1055–1074.

Janzen, M. E., & Vicente, K. J. (1998). Attention allocation within the abstraction hierarchy. *International Journal of Human-Computer Studies, 48,* 521–545.

Kelley, C. R. (1968). *Manual and automatic control.* New York: Wiley.

Law, A. M., & Kelton, W. D. (1991). *Simulation modeling and analysis* (2nd ed.). New York: McGraw-Hill.

Leveson, N. G. (1995). *Safeware: System safety and computers.* Reading, MA: Addison-Wesley.

Maddox, M. E. (1996). Critique of "A longitudinal study of the effects of ecological interface design on skill acquisition" by Christoffersen, Hunter, and Vicente. *Human Factors, 38,* 542–545.

Pawlak, W. S., & Vicente, K. J. (1996). Inducing effective operator control through ecological interface design. *International Journal of Human-Computer Studies, 44,* 653–688.

Reising, D. C. (1999). *The impact of instrumentation location and reliability on the performance of operators using an ecological interface for process control.* Unpublished Ph.D. thesis, University of Illinois at Urbana-Champaign.

Reising, D. C., & Sanderson, P. M. (1996). Work domain analysis of a pasteurization plant: Building an abstraction hierarchy representation. In *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting* (pp. 293–297). Santa Monica, CA: Human Factors and Ergonomics Society.

Reising, D. C., & Sanderson, P. M. (2002a). Ecological interface design for Pasteurizer II: A process description of semantic mapping. *Human Factors, 44,* 222–247.

Reising, D. C., & Sanderson, P. (2002b). Work domain analysis and sensors: I. Principles and simple example. *International Journal of Human-Computer Studies, 56,* 569–596.

Reising, D. C., & Sanderson, P. (2002c). Work domain analysis and sensors: II. Pasteurizer II case study. *International Journal of Human-Computer Studies, 56,* 597–637.

Sakuma, A., Itoh, J., Yoshikawa, E., & Monta, K. (1995). Simulation study of an ecological interface for nuclear power plants. In Y. Anzai, K. Ogawa, & H. Mori, (Eds.), *Symbiosis of human and artifact* (pp. 989–994). Amsterdam: Elsevier.

Vicente, K. J. (2002). Ecological interface design: Progress and challenges. *Human Factors, 44,* 62–78.

Vicente, K. J., Moray, N., Lee, J. D., Rasmussen, J., Jones, B. G., Brock, R., et al. (1996). Evaluation of a Rankine cycle display for nuclear power plant monitoring and diagnosis. *Human Factors, 38,* 506–521.

Vicente, K. J., & Rasmussen, J. (1992). Ecological interface design: Theoretical foundations. *IEEE Transactions on Systems, Man, and Cybernetics, 22,* 589–606.

Yamaguchi, Y., & Tanabe, F. (2000). Creation of interface system for nuclear reactor operation: Practical implication of implementing EID concept on large complex system. In *Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Society* (Vol. 1, pp. 571–574). Santa Monica, CA: Human Factors and Ergonomics Society.

Dal Vernon C. Reising is principal research scientist at the Knowledge Services Laboratory, ACS Advanced Technology, Honeywell International, Inc., in Minneapolis, Minnesota. He received his Ph.D. in industrial engineering in 1999 at the University of Illinois at Urbana-Champaign.

Penelope M. Sanderson is a professor in the Key Center for Human Factors and Applied Cognitive Psychology at the University of Queensland. She received her Ph.D. in psychology in 1985 at the University of Toronto.

Date received: September 1, 2002
Date accepted: December 30, 2003