Beyond Ecological Interface Design: Lessons From Concerns and Misconceptions
Beyond Ecological Interface Design: Lessons from Concerns and Misconceptions

Clark Borst, John M. Flach, Joost Ellerbroek, Member, IEEE

Abstract—The Ecological Interface Design (EID) paradigm was introduced in the process control domain twenty-five years ago by Kim Vicente and Jens Rasmussen, as a way to help operators cope with system complexity and events unanticipated in the design of automated control systems. Since that time, this perspective has sparked interest in other safety-critical, sociotechnical domains where humans cooperate with computerized systems to ensure safe and efficient system behavior. Many of our own, but also other explorations have, however, resulted in several usability concerns and misconceptions about the EID perspective as a viable design approach. This paper discusses some of these concerns and misconceptions, where the final goal is to get past the EID label and to consider the general lessons relative to the demands and opportunities that advanced information technologies offer, and complex systems require. The article concludes with a preliminary outlook for the future of EID, where it is anticipated that the adjective ‘ecological’ will become increasingly redundant, as the focus on supporting ‘productive thinking’ becomes the dominant paradigm for engineering representations.

Index Terms—Ecological Interface Design (EID), Cognitive Systems Engineering (CSE), Aviation, Productive thinking

All new technologies develop within the background of a tacit understanding of human nature and human work. The use of technology in turn leads to fundamental changes in what we do, and ultimately in what it is to be human. We encounter the deep questions of design when we recognize that in designing tools we are designing ways of being. By confronting these questions directly, we can develop a new background for understanding computer technology – one that can lead to important advances in the design and use of computer systems. (Winograd & Flores, 1986, [1, p. x])

I. INTRODUCTION

As suggested by the opening quote, advances in information technologies (e.g., automatic control systems, sensors, displays, etc.) have changed how we think about work, how we think about human cognition, and how we think about the design of computer systems. While the role of humans in these systems has changed, it has not been diminished [2], [3]. Over the years, particularly in the aviation domain, it has become evident that humans continue to be an essential component in technical systems, as they can bring adaptivity and creativity that can enhance the resilience of sociotechnical systems. Such abilities are highlighted when a human performs “heroic” acts and recoveries, as in the Hudson River landing where Captain Sullenberger successfully landed his Airbus A320 on the Hudson River after losing thrust in both engines due to multiple bird strikes [4].

Therefore, rather than striving exclusively to replace human weaknesses with technical systems, the emphasis has shifted to exploring ways in which technology can facilitate human adaptivity and flexibility to cope with unforeseen events (i.e., to enhance resilience). In complex domains there will always be a potential for problems that cannot be anticipated in the design of automatic control systems. The creative human expert therefore remains an important resource for dealing with this unanticipated variability. The key implication for the design of interfaces and automation systems is that cooperation between automated control systems, decision support systems and humans is required for the system to respond robustly to complex work environments [5], [6]. This coordination between humans and automatic systems requires interface representations that are grounded in constraints that reflect the deep structure of the work domain demands [7], [8].

Recognizing these changes in the human role, Rasmussen and Vicente introduced the concept of Ecological Interface Design (EID) in the process control domain twenty-five years ago [9], and elaborated further on this concept in two successive papers [10], [11]. In contrast to user- and technology-centered approaches that put the emphasis on either the human or the technology, EID starts by focusing on the work domain (i.e., ‘ecology’) as this facilitates and bounds purposeful patterns of activity, irrespective of human or automated agents that perform the activities. The goal of EID is then to facilitate coordination between humans and automatic systems by making interface representations that reflect the deep structure of the work domain in ways that support human skill-, rule-, and knowledge-based problem-solving activities, for example, through metaphorical design and direct manipulation [11]. Since that time, books [8], [12], [13] and numerous papers have explored the EID approach for the design of graphical interfaces and decision aids for controlling complex systems in various application domains, such as process control [14], health care [15], command and control [16], and aviation [17]–[24]. As with any new construct, these explorations have generated both signal and noise, and have occasionally resulted in usability concerns and misconceptions about the perspective.

This paper discusses some of these concerns that have been articulated in the literature, or in some cases reflect the initial skepticism that guided our own research to explore the
boundaries of the EID approach in the aviation domain—a domain that is increasingly more focused on optimization [25], [26], automation with increased control authority [27], analytical approaches to dealing with sensor failures [28], and hiding of information to reduce the apparent complexity of the control problem and manage workload [3]. This paper will therefore be organized around the following five concerns and/or misconceptions about the EID approach that we have encountered over the years:

1) Interfaces designed according to EID principles will be “natural” and “simple” to use, requiring minimal training.
2) Interfaces designed according to EID principles will be vulnerable to catastrophic failure if there is a sensor failure.
3) Interfaces designed according to EID principles will lead to inefficient solutions for dynamic control problems.
4) Interfaces designed according to EID principles will invite people to “push the envelope”, leading to unsafe system performance.
5) Interfaces designed according to EID principles will become less important as information technologies improve and take over more and more of the human’s control activities.

These concerns are articulated in terms of ‘EID principles’ in recognition of the important role that Rasmussen and Vicente’s papers had in inspiring our work to improve interfaces in the aviation domain. The goal of this paper, however, is to get past the EID label, to instead consider the general lessons relative to the demands and opportunities that advanced information technologies offer. The principles addressed go beyond the EID approach, and the lessons learned through EID-inspired work are relevant to all who are interested in supporting productive thinking through interface design.

II. EID IS NOT SIMPLE/NATURAL: THE LAW OF REQUISITE VARIETY

A dominant misconception about EID is tied directly to the choice of the term “ecological,” which is often interpreted as “natural” and “intuitive.” For example, Oshima and Wickens mention that ecological displays would have a natural presentation of task-relevant information, leading to relatively automatic processing of information [29]. However, they were referring to ecological spatial displays that enhance optical flow, a cue humans use for the perception of ego-motion in their natural environment, instead of the supervisory control of complex, automated systems. In another report, Chen et al. state that “EID is a user interface design technique that conveys the constraints in the tasking environment, usually visually via emergent patterns, so the operator can intuitively perceive and solve the problem” [30, p. 22]. Again, ecological displays seem to be tied to intuitive displays. This in turn can lead to a pervasive misconception, especially in the aviation domain, that an ecological interface will be natural, intuitive, and therefore easy to use.

Given the Oxford dictionary definition of intuition as the ability to acquire knowledge without inference or the use of reason, there seems to be an expectation that with an EID interface, someone with little or no training will immediately be capable of expert performance in controlling a complex process. Vicente, however, always intended EID as a method to develop displays for domain experts, which has also been the conclusion of several evaluation studies. That is, displays designed according to EID principles are often very complex, and typically require extensive experience before people are capable of taking advantage of the power of these representations [31], [32]. Results from several follow-up studies suggest that domain experts are usually the ones that can pick up an ecological display more easily than non-experts [33], [34]. Despite being criticized [35], empirical evidence also exists that prolonged working experience with ecological interfaces can actually help operators develop skills and expertise [31], [32]. While this does show that ecological displays are or can become intuitive to expert operators, it can not be taken to mean that ecological equals intuitive by default.

Actually, the term “ecological” was chosen in recognition of the work done by Brunswik and Gibson [9], [36], [37]. Rasmussen and Vicente were inspired by the fact that both Brunswik and Gibson emphasized the significant role of the ‘stimulus’ or ‘ecology’ in shaping human performance. This is perhaps best represented in Brunswik’s insistence on representative experimental design, and in Gibson’s construct of affordances [36]–[38]. In the context of complex human-machine systems, however, the fundamental principle that reflects EID’s orientation to the ecology is Ashby’s Law of Requisite Variety [39], [40].

In its most succinct form, the Law of Requisite Variety states that “only variety can destroy variety.” To put this in the context of EID, the point is that for any representation to be effective for communication and/or control, it must be as complex as the problem to be solved. In other words, an effective representation must not trivialize the problem. A good representation’s ‘variety’ or ‘complexity’ must be as ‘rich’ as the problem that needs to be ‘solved,’ ‘controlled,’ or ‘destroyed.’

The Law of Requisite Variety is grounded in Shannon and Weaver’s theory of communication [41, p. 8–9]. They state that ‘information’ relates not so much to what you do say, as to what you could say. That is, information is a measure of one’s freedom of choice when one selects a message.” Thus, in representing a problem space or a work domain, the requisite variety reflects the space of possibilities, the space of what can be done. The more fully this space is represented, the wider the capacity of the resulting control system will be to move within that problem space. It can therefore be argued that the first priority of the EID approach is to ensure that the representations developed completely and accurately depict the space of possibilities.

There are two ways that a representation can fail with regards to this prescription. On the one hand, a representation can trivialize a problem in a way that greatly under-represents the full space of possibilities. This is particularly likely when the focus of design is on the limitations of human information processing. This can result in integral or configural displays that are designed around a narrow range of possibilities,
typically associated with the behavioral task requirements and the normal operating ranges of a system [8]. While such displays might be effective for control under the limits of normal operating assumptions, often these displays fall short as decision aids for fault diagnosis, especially in response to situations that were not or could not have been anticipated at the design phase.

A second way in which a representation can fail to meet the prescription of EID is to make the problem more complicated than it needs to be. This is particularly likely under the single-sensor-single-indicator (SSSI) philosophy [42] that drove many early designs in nuclear power systems, and to some extent in early but also still in modern aviation displays (Fig. 1). To understand the flaw in the SSSI philosophy, it is useful to make an analogy between the sensors in a system like a nuclear power plant and the alphabet of a language like English. In design of a high-risk system, engineers are generally well aware of the constraints of requisite variety, and they generally try to ensure that every important dimension of the plant (i.e., every letter of the effective alphabet) is sensed or measured. In turn, in the design of the control room for the plant, each of the sensed variables is represented by a single dedicated display. It should not be a surprise to anyone that such designs lead to cognitive overload in emergency situations, such as a loss of coolant or hydraulic power, where many sensors and associated alarms will be flashing simultaneously. Does this suggest perhaps that the requisite variety of a power plant exceeds the information capacity of humans – thus, requiring a shift from human to control by computers whose capacity to destroy variety might greatly exceed that of humans?

This is certainly a legitimate question and many have concluded that this is indeed the case. EID, however, suggests another path. There is something missing in the SSSI representation that is fundamental to EID, communication theory, and requisite variety. When an alphabet of sensors (e.g., consider 26 letters) is represented in the SSSI format, there is an implication that at any moment every possible configuration is possible. However, as we know with natural languages there are significant constraints on what letter combinations are possible. Certain letters like ‘e’ and certain combinations like ‘th’ are very frequent, while other letters like ‘x’ and other combinations like ‘bd’ are rare. Because of these constraints, it is possible to read a text even when many of the letters are occluded. Similarly, in a power plant, the possible outputs of various sensors are often quite heavily constrained by physical laws such as mass and energy balances, so that there are far fewer possibilities than suggested by the SSSI format.

This insight, that there is a deep structure (i.e., constraints) that limits the ultimate complexity of a problem (i.e., the requisite variety), was fundamental to Wertheimer’s observations about the impact of alternative representations on ‘productive thinking’ [43]. Here, the term ‘productive thinking’ reflects early work on problem solving by Gestalt Psychologists. In Wertheimer’s book, he provides many examples that illustrate how the difficulty of a problem can be dramatically impacted by the nature of the representation, e.g., a problem that is difficult with one representation can be simple with another representation. Thus, productive thinking reflects a relationship between the organization in a representation (either internal or external) and the structure of a problem (i.e., constraints). EID balances the demands of requisite variety against the limited information capacity of humans by making the operating constraints (e.g., physical laws) explicit in the representation. Organizing or configuring information to reflect constraints associated with the work processes will typically suggest meaningful ways for operators to ‘chunk’ information, reducing the demands on memory and supporting productive thinking.

A good example is the EID interface for DURESS, which uses a funnel metaphor to make the constraints associated with mass and energy balances, captured in the first law of thermodynamics, explicit and easy to understand [12]. Similarly, Borst et al. showed that, by making aircraft climb performance and energy state explicit on a Synthetic Vision Display, pilots are able to make meaningful assertions about
the aircraft’s capability to safely fly over terrain, and the various ways to do so [22]. Fig. 2 illustrates the display proposed by Borst et al., and shows some of its meaningful properties. For instance, the display not only shows that a collision with terrain is imminent (the displayed flight-path vector points into the terrain), it also shows that the aircraft’s climb performance allows it to climb over the terrain (the optimal straight and turning climb angles lie above the terrain), as well as the relationship between the total energy (kinetic and potential) and the desired climb angle. Portraying such aircraft performance constraints in a way that allows pilots to visually relate them to external terrain constraints can reduce the cognitive load from knowledge-based assertions about the aircraft’s maneuverability to rule-based shortcuts in the form of “if-then” conditionals. When a pilot makes the decision to clear the terrain by climbing, skill-based manipulation of the elevator and throttle is used to align the flight-path vector and the total energy angle with the desired climb angle to reach a stable and sustained climb state. More laborious knowledge-based problem solving would be triggered when the aircraft’s climb state cannot reach the expected nominal sustained climb.

To summarize, the goal of ecological representations is not to trivialize complex work such that an ecological interface will be natural and easy to use for someone with little or no experience, but rather to organize the information in ways that leverage the deep structure of the work domain (i.e., constraints). The impact of this approach is that operators can learn to ‘see’ the data in the context of that deep structure, so that meaningful associations among the variables are more salient. For complex work, this will not eliminate the need for learning and training, but it should facilitate developing the types of associations that will eventually support skill-, rule-, and knowledge-based solutions to complex problems. Thus, ecological displays are not a substitute for human expertise. Rather, they should be designed to complement that expertise by representing the deep structure that experts need to think productively. Note, that this is equally relevant for any interface design method.

III. EID AND SENSOR FAILURE VULNERABILITY

Vulnerability to sensor failures is perhaps one of the most well-documented concern about ecological displays. It was initially pointed out by Vicente and Rasmussen in 1992 [11], reiterated upon four years later [44], and later again cited as a potential ‘showstopper’ [45]. The gist of the concern is that people may continue to trust an EID display even when the information driving it is unreliable (e.g., as the result of a failed or noisy sensor). Thus, essentially, the fear is that the principles of EID that amplify the signal will be equally powerful in amplifying noise.

If one thinks carefully about communication theory, one can easily see the flaw in this argument. Imagine a failed key on a typewriter or keyboard (perhaps that left a blank whenever the letter ‘e’ was typed). Would this failure be easier to discover in a written text like a novel (where the constraints of the language are explicit) or in a random list of letters with no explicit constraints? Of course, when the constraints were more explicit, the key failure will also be made more explicit, not less so.

Similarly, when the constraints governing a process are made explicit in a representation, then whenever a sensor behaves in an incorrect way that behavior will be far more salient than when the behavior is presented in a context-free form. The Turkish airline accident near Amsterdam Schiphol on February 2009 [46] illustrates the danger of presenting information in a context-free form. There, a radar altimeter failure made the sensor readouts behave in an arbitrary way. Although the pilots were aware of this faulty sensor, they did not realize that the auto-throttle system was using it to regulate the engine thrust in the approach phase. As a result, both engines were commanded to ‘idle’ at 2,000ft, which made the airplane lose too much altitude and speed. After reaching a stall at 460ft, the aircraft eventually crash-landed short of the runway. The important lessons that can be learned from the Turkish airline accident are not only that pilots can sometimes have incomplete, or even incorrect, mental models of automated flight controls, but also that pilots are not always able to put isolated sensor failures and their implications into context.

In aviation, sensor failure diagnosis is commonly approached by analytical and model-based methods [47]. This focus on ‘automatic’ techniques suggests that sensor failures and their implications are not something to be left for humans to deal with. After all, the faulty radar altimeter readouts in the Turkish Airlines flight were visible on a display, but the pilots failed to take appropriate action accordingly [46]. However, it can also be argued that if the information flow from sensor to
auto-throttle system was made clear, and if the display showed the predicted touchdown point based on the sensed information and the autopilot’s intent, the fault would have been more salient. What would be the role of the EID perspective in such solutions?

The EID framework suggests a constraint-based approach to expose interactions between system components and system functionalities. In effect, the constraints form an explicit model of expected behavior and violations of expectancy due either to sensor failure or any other system fault will become salient as deviations from the model. In most systems, it is likely that sensor failure will have a signature that is distinct from other types of faults. This will be due to how the failures propagate through the system. Sensor failures will generally be distinctive in that they will be uniquely isolated to a single sensor and should not propagate in the same way that a physical fault would. Additionally, by ensuring one-to-one mappings between work domain constraints (at different levels of abstraction) and interface cues, functional redundancy would be added that is similar to the concept of analytical redundancy found in control theory [48].

However, it is not necessary to rely solely on logic or theory. A number of studies have explicitly examined EID under contexts of sensor and system failures, and the consistent result is as implied by the logic of communication theory. For example, Borst et al. showed that an aircraft engine failure or jammed flaps became more salient in a constraint-based Synthetic Vision Display (SVD), which displays more higher-order functional details of the underlying work domain structure, than with the baseline SVD that only showed lower-order work domain information [22]. In particular, the discrepancy between nominal and actual sustained climb performance made the pilots aware that something was amiss, enabled them to trace back the cause of the discrepancy, and triggered knowledge-based problem-solving activities to find alternative escape maneuvers.

In the automotive domain, a study on the effect of visualizing adaptive cruise control limits showed that their EID display enhanced driver response in both the manual control of car following and in responding to cruise control failures [49]. In that study, it was concluded that “the EID provided a continuous indication of automation and roadway state that helped to make the subtle cues associated with braking behavior of the lead vehicle more salient, prompting appropriate reliance in terms of the detection and response to automation failure when braking limits were exceeded.” [49, p. 204]

Studies with DURESS III that examined the effect of sensor failures suggested that the performance and control stability of the EID participants were not significantly compromised by increased sensor noise. Moreover, it also appeared that performance and control stability with the ecological interface were consistently superior to that of the SSSI representation [50], [51].

In the process control micro world of the Pasteurizer II, Reising and Sanderson showed that sensor failures became more salient and easier to detect when using an ecological display with maximally adequate instrumentation than when using an equivalent piping-and-instrumentation interface [52]. Their results also indicated that failure diagnosis deteriorated significantly with an interface that showed only a subset of constraints and relationships. Reising and Sanderson concluded that if ecological interfaces are to be effective for failure diagnosis, “they should be free of ambiguities that arise when their sensors are only minimally adequate to derive and display higher-order system properties.” [52, p. 332] In other words, for effective control and fault diagnosis it is crucial to ensure that all information that should be displayed will indeed be displayed. Note that this prescription is also well known in control theory, in the concepts of ‘observability’ and ‘controllability’ of a system, and that it therefore equally applies to any interface design method.

To summarize, when the constraints of a control problem are made explicit, violations of the natural constraints (i.e., an isolated sensor giving readings that are not constrained by the physics of the process) are also made more explicit, not less explicit. This helps operators to identify and diagnose sensor failures as they become deviations from an explicit model of expected behavior. This does, however, depend on the quality of the mapping of the work constraints to the interface. If significant properties of the deep structure are not represented in the interface, then there will be ambiguities (including sensor failures) that may be difficult for operators to detect and resolve.

**Fig. 3: An example of a constrained action state-space, with an optimal trajectory, as well as several (sub-optimal) alternatives.**

**IV. EID AND CONTROL TASK EFFICIENCY**

EID has an explicit appreciation for the flexibility in which control problems can be solved. Whereas more classical approaches to systems and interface design seek to optimize behavior, and look for particular ways of performing a task, EID tends to focus on ‘control structure’ instead of ‘control behavior’ [11]. This implies that operators are free to choose any strategy they prefer, as long as it does not violate work domain constraints (Fig. 3). This, however, also implies that operators can choose suboptimal strategies that may fail to comply with, for example, the efficiency goals of a system. Given that current developments in aviation, such as the design of the next generation air traffic management system [25], [26], are specifically focused on ‘optimization’ (i.e., optimal landing sequences, optimal fuel usage, and optimized flight trajectories), ecological information systems may not be
considered suitable for control problems that benefit from a particular control strategy that seeks for the optimized and best possible solution.

This concern therefore reflects the potential conflict between more classical approaches to control system design and the EID approach, in terms of the attitude toward solutions to work problems. Control engineers are often trained to seek the ‘one best way’ (i.e., the optimal control solution [53]). While this may be feasible for well-defined problems, it is not always desirable for complex work domains. In complex domains, it may be necessary to shift the focus from designing control solutions that minimize a particular control situation (e.g., specific process model or cost function), to designing solutions that can satisfy a wide range of potential situations. In other words, the focus shifts from optimal control to robust control [54], [55]. Similarly, EID is designed to help operators choose from a range of satisfactory strategies, and facilitates a shift in strategy when warranted by the changing demands of a process. The rationale behind this is that in reality many factors are and will remain unknown (such as local adverse weather in aviation) and unexpected disturbances may force the user to deviate from an optimal plan to ensure safety.

It is therefore not surprising that, when ecological systems are deployed in operational environments that are (historically) organized from a centralized control perspective, operators often adopt strategies that satisfice instead of seeking an optimal solution. Rather than concluding that EID is unfit for such control problems, it can be argued instead that new metrics to describe the control problem are required, ones that describe the higher order constraints across multiple control dimensions and the value propositions that may be contingent on situational variables. Because common performance metrics in most engineering models assess deviations from an optimal solution, it is not surprising that under some conditions the ecological information aids will perform less well than systems specifically designed to follow the optimal strategies. It is, however, also true that optimal control solutions may fail catastrophically when the problem changes in ways that violate assumptions of the analytic model. Note that this critique is typically true for all control systems that are designed to be robust over a wide range of situations. In these cases, to ensure this robustness, it is generally necessary to sacrifice optimality relative to any particular situation. Thus, a robust or resilient control system will typically be designed to satisfice. That is, it will not necessarily be optimal for any particular situation, but it will be good enough (i.e., perform within some tolerance range) over a large variety of situations.

This difference between a desire for robustness and optimality is also reflected in the difference between (cognitive) work analyses and (cognitive) task analyses, a difference that is also sensitive to misinterpretation. From the perspective of the work analyst, it is important to differentiate between work domain constraints, and the constraints associated with particular solutions or strategies for navigating those constraint spaces (i.e., the constraints associated with a particular control solution) [12]. Any particular control solution typically introduces additional constraints that will shape the control activities. For example, a stationary model of plant dynamics or a fixed ‘cost functional’ reflects constraints of particular control solutions that may not be representative of the full complexity of the actual work domain, where plant dynamics or values associated with competing goals may change. Such limitations of automatic control solutions are one of the reasons that supervision by human operators remains critical in complex work domains. Note also that such limitations of particular control solutions apply to any fixed procedure, whether implemented by automation or human.

For the design of adaptive control solutions, however, it is critical to differentiate between the constraints associated with any particular control solution and the constraints associated with the deep structure of the work ecology. This is a primary goal in designing EID representations: To allow operators to evaluate particular control solutions in the context of the deep structure of the work domain. This is necessary for the operators to be able to intervene (e.g., take over from the automation or change their own procedures), when changes in the work domain require a shift of control tactics or strategy in order to satisfy demands. Thus, the concern that EID is not suitable for finding efficient solutions to dynamical control problems is grounded in a failure to understand the difference between control theory (i.e., the search for good solutions to control problems) and particular control solutions. Any particular control solution will be ‘bounded’ by the assumptions of its designers. But control theory, like cognitive work analysis, provides a framework for considering the range of possible solutions and for attempts to map the full requisite variety of a problem, before selecting any particular solution. This means that control solutions are constantly being evaluated and improved as designers continue to explore the requisite variety of a problem.

A good example of a distinction between a particular control solution and the deep structure in the work domain can be found in the design of current autopilots. Commonly, autopilot control laws are created on the basis of ‘small-perturbation’ engineering models of the airplane dynamics [56]. For example, in the control laws for vertical flight-path navigation, altitude and aircraft speed are chosen to be the parameters (or states) that need to be controlled [57]. The control strategy for regulating altitude and speed typically consist of a pitch angle autopilot (controlling the elevator control surface) and an auto-throttle that both work independently. Such an architecture creates the illusion that an aircraft can assume any thinkable combination of speed and pitch orientation. The same illusion may also be preserved (or even amplified) on the Primary Flight Display that shows speed and altitude on separate regions on the display. This is of course not the case at all. Elevator inputs not only affect the pitch angle, but also speed. Vice versa, throttle inputs not only affect speed, but also pitch orientation. So speed, pitch orientation, and altitude are not independent at all, but coupled. The classical control strategy applied in autopilots does not take advantage of constraints associated with this coupling, and may therefore be less agile, especially in the face of unanticipated system variabilities.

What is missing from the small-perturbation control models is the possible exchange of potential (altitude) and kinetic (speed) energy, which relies on the natural law of energy
concentration. A combination of an autopilot and auto-throttle design that does recognize the importance of energy management is the Total Energy Control System (TECS) by Lambregts [58], [59]. Here, the throttle is used to control the sum of kinetic (speed) and potential (altitude) energy, while the elevator is used to control the balance between these. TECS yields a more natural airplane behavior by incorporating energy management in the design.

To summarize, the concern regarding EID’s suitability for control problems that benefit from particular (optimal) control strategies reflects the potential conflict between more classical approaches to control system design and the EID approach. Whereas classical engineering approaches work toward finding a particular control strategy that reflects the most efficient solution to a particular situation, EID strives to work toward more robust solutions that will be good enough over a large variety of situations. In that sense, the concern is legitimate in that EID may sacrifice optimality relative to any particular situation. However, for complex work, where system dynamics or values associated with competing goals may change, resilient solutions will generally be preferred to solutions that are optimal most of the time, but that can fail catastrophically for some small set of situations [60].

V. EID AND SAFETY: MIGRATION TO ACCIDENTS?

During our explorations of the EID perspective in aviation, we observed that pilots were sometimes migrating to the limits of safe system performance. That is, we consistently found in various experiments that pilots sometimes undertook risky actions with ecological displays, by maneuvering themselves in narrow control spaces that leave little room for error [21], [22], [61]. This gave rise to the concern that ecological information systems could invite risky control actions.

The undesired tendency of humans to seek out the limits of safe or acceptable operation was already mentioned by Rasmussen in his model of ‘Migration to Accidents’ [62]. In this model he describes how the pressures for efficiency, and a tendency for the least amount of effort, can cause the operator to systematically migrate closer to the limits of system performance. With ecological displays, operators can directly perceive these limits and thus may be more likely to ‘push the envelope’ than operators who have more uncertainty about the location of such a boundary. Are ecological information systems therefore unsafe?

Although we initially raised the concern regarding the tendency of operators to seek out the limits of system performance, we do not believe that this implies that ecological information systems are unsafe by default. Note, that the likelihood of migration to accidents as described by Rasmussen is much higher when the bounds on safe operation are not well specified (e.g., they are not discovered until they are crossed). One of the original motivations of the EID approach was to provide an information counter weight to the inevitable pressures for increased efficiencies that lead to ‘normal accidents’ [64]. We argue, instead, that the source of the undesired system performance is not tied to the EID approach itself, but rather to the scope of the work domain that has been modeled. Most work domains contain a mixture of causal (physical) and intentional constraints (i.e., organization, procedures, rules, and laws) that jointly contribute to establishing the conditions for safe operation (Fig. 4). For example, aviation safety is not only accomplished by the technical systems onboard an aircraft, but also by standardized communication and coordination protocols, standard procedures, and standards for organizing the airspace. These non-technical aspects of safety are sometimes referred to as ‘safety culture.’

The EID approach can be used to make both the physical and the intentional constraints visible, and it can also manipulate the relative salience of those constraints. By making intentional constraints salient it will typically make the system ‘safer’, but will have potentially negative consequences in terms of efficiency and robustness. That is, it can bias operators to not take advantage of the full field of possibilities set by the physical constraints, and it may bias operators against potential solutions to exceptional situations that demand violation of intentional constraints to achieve stability. A good illustration of this conflict is the flight envelope protection system that is present in several modern commercial aircraft. Such a system would qualify as a purely intentional ‘safety’ system that prevents an aircraft from entering unsafe flight regimes (by a margin) and unusual attitudes that are difficult to recover from. In the context of terrain avoidance, however, such a system may be too restrictive, when a flight into terrain needs to be averted. For example, in the China Airlines B-747 incident in 1985, the flight crew was forced to overstress (and structurally damage) the horizontal tail to recover from a rapid vertical dive that plunged the aircraft 30,000 feet from an original high-altitude cruise [65]. It is very likely that the aircraft would have crashed if an envelope protection system had been present that prevented the pilots from initiating control commands that would lead to structural damage.

In general, the scope of the work domain analysis determines the accepted safety level reflected by ecological systems. On the one hand, when the scope is too much on the causal constraints, the ‘physical structure’ in the environment will be made compelling and this can cause people to pursue these physical boundaries, leaving little room to prevent accidents. This was also the behavior that was observed in the aviation examples. On the other hand, when the scope is too much on the intentional constraints, the operational range of physical
systems can be too limited to effectively solve problems in exceptional situations. The challenge is therefore to find the right balance between causal and intentional constraints, and their representations in ecological information aids.

In a sense, the relative salience and balance between physical and intentional constraints is actually a matter of how much we trust the operators to participate in the decisions associated with efficiency and risk tolerance. If we do not trust the operators, then the intentional constraints should be made very salient or we may actually limit the action capability of the operators to prevent actions that cross the intentional boundaries. However, such a system will be brittle in the face of unanticipated variability. If we do trust the operators, then they should have as much information as possible so that they can make smart decisions with regards to trading off safety and efficiency. The EID approach is often motivated by the assumption that you can trust the operators’ expertise, because they will typically have information that was not available to those making the original choices with regards to changing dynamics that impact risk and efficiency.

Preliminary empirical evidence has been demonstrated by Comans et al. [66]. They investigated the effects of showing intentional constraints in conjunction with physical constraints on pilot’s compliance with minimum safe altitude rules and decision-making strategies in an aircraft terrain clearance task. It was found that when intentional constraints were visualized in addition to physical constraints, pilots not only migrated closer toward the intentional system boundary, but also opted for less risky control strategies to clear terrain as compared to the visual representation of only physical constraints. Additionally, the variability in control behavior between pilots also decreased, resulting in more uniform control actions and terrain clearance buffers that better complied with the intentional system boundaries.

To summarize, the concern that ecological displays are unsafe, because operators tend to seek out the limits of system performance, is directly tied to the scope of the analysis, rather than to the EID approach itself. For many complex socio-technical systems, the intentional structure in the work domain is often overlooked, and is not taken into account in the work analysis and display design. As a result, the physical structure will be made compelling, which can invite operators to push the physical envelope. By also representing intentional system boundaries, it has been shown that operators can make smarter decisions with regards to the tradeoff between safety and efficiency. However, at the end of the day, the EID approach does not dictate a particular balance between physical and intentional constraints. The goal for the EID approach is not to dictate a specific safety solution in the design of the interface, but to give the operator the information about both intentional and physical constraints so that they can make safe choices in situations that could not have been anticipated at the design stage.

VI. EID AND AUTOMATION: CLOSING OR OPENING THE BLACK BOX?

One of the characteristics of complex sociotechnical work domains is the relatively high degree of automation. Especially aviation has a long history in automation, where the first (mechanical) autopilot was developed and demonstrated ten years after the first successful flight by the Wright Brothers in 1903 [3]. Since then, the number of automated systems has increased dramatically, and pushed the flight crew into the role of system supervisor and computer manager. Because the majority of the ecological interfaces that have been developed over the last twenty-five years mainly support manual control tasks, where the computer is used for information integration and acquisition to compose the visual representation, an impression can be created that ecological interfaces are solving an old problem. As such, a misconception can arise that there is no need for interfaces that reveal the deeper structure of the work domain when automation is taking over more and more of the control activities.

In aviation, for instance, the dark-cockpit concept [67] implicitly points to this misconception. In this concept, information will not be displayed until something goes wrong. In other words, an annunciator panel light remains dark until the pilot needs to be notified of an abnormal condition. In future cockpit prototypes (e.g., Thales’ Cockpit 3.0), where the entire instrument panel has been replaced by touch screen technology, this concept is taken one step further. Here, almost the entire cockpit stays dark and only the information is displayed that the automation considers to be relevant for a particular phase of flight. Despite the fact that such a concept may reduce workload and simplify pilot scanning patterns, there is a clear ambiguity in the message the dark cockpit concept delivers. That is, a dark cockpit can either be an indication that everything is normal, yet it can also be the result of a failure in the display or the software driving it, thus potentially hiding a serious problem.

System designers often have valid reasons (and sometimes an over-tendency) to increase the level of automation for safety, efficiency, and workload management purposes. Woods and others make clear that increasing the level of automation is not good or bad in itself, but that with more automation greater coordination between people and technology will be required [68]–[71]. They argue that with increased automation, a rich information coupling becomes even more important to ensure human-machine coordination. That is, in order to cooperate with smart technologies you actually need more information (i.e., richer interfaces), not less. This coordination becomes especially important when humans are expected to maintain final control authority, and therefore must be able to intervene in situations unanticipated in the design of automated systems. To mitigate breakdowns in coordination and their implications, it is imperative to create new tools for coordination, so as to make automated systems ‘team players’ [72].

When looking at human-human interaction, productive team thinking and problem-solving efforts are accomplished when teammates have a “common ground” or shared representation of understanding of the work to be done and the various ways to do it [73]. Similarly, when work is distributed over human and automated agents, the constraints introduced by the other agents are properties of the work domain or problem space. These constraints must be considered in order to achieve productive team thinking. Thus the requirement for coordination
Fig. 5: When the rationale that guides the machine is based on the constraints and relationships of the work domain, then the interface revealing them becomes the operator’s window to both the work domain itself (required to solve a control problem manually) and the machine’s rationale (required to monitor the machine). Additionally, the constraints introduced by the cooperating agents are properties of the work domain and must therefore be considered in the design.

among multiple agents is part of the deep structure of the control problem, and these demands have to be included in the representation of the work. This requires that both the design of the automation and the design of the interface are guided by an analysis of the deep structure (semantics) of the work or problem domain.

Before making any decision with regards to the design of a particular human-machine system, it must be taken into consideration that automated agents can only be team players when they are both observable and directable [70]. That is, operators need to be able to understand and assess the rationale that guides the automation, and operators must be able to intervene and re-direct machine activities when warranted by the demands of the situation (Fig. 5). Therefore, instead of just making more (low-level) data about the machine’s activities available to the user, it is important to find a representation that maps the problem space in such a way that it utilizes the natural constraints most effectively.

The key challenge in designing such a shared representation is to find the right state variables or coordinate system for thinking productively about the problem. In EID, this is done by amplifying system invariants, and by utilizing metaphors that reflect the control problem to leverage the experience of the interface users [8]. In addition it must be ensured that this coordinate system is also the common ground underlying the coordination between humans and automation. As such, the interface will become a window to both the rationale that guides the automation (needed to monitor and observe the automation) as well as to the laws and rules governing the work domain (needed to solve a control problem manually) [71]. An example of such a shared representation can be found in an airborne separation assistance display proposed by Van Dam et al. [20] and Ellerbroek et al. [23], [24]. Whereas the critical control variables in most airborne separation assistance and collision avoidance systems are based on closest-point-of-approach models, which require look-ahead flight-path predictions and condition checks to detect safety hazards, this alternative representation relies on the underlying and more natural principle of relative navigation between moving vehicles.

The velocity obstacle theory, commonly found in robotics, provides a formal domain representation that captures the underlying principle of relative navigation between moving vehicles. For a moving obstacle and a given separation margin, the collection of conflicting relative velocities (or obstacle lines) forms a visually attractive triangular area, called a velocity obstacle [75]–[77]. In the display concept proposed by Van Dam et al., this velocity obstacle is directly portrayed on a pilot’s navigation display (Fig. 6) to support the pilot’s ability to visually detect conflicts and to simultaneously observe the complete space of maneuver possibilities (in both heading and speed) to resolve (multiple) conflicts in the horizontal plane [20]. Additionally, the velocity obstacle representation can also serve as the underlying rationale that guides conflict resolution automation to ensure coordination between human and automated agents. In other words, utilizing the velocity obstacle representation both in the design of the interface and in that of the automated agent not only enables the human agent (e.g., a pilot) to observe the decision-making criteria of the automated agent, but also provides the means for the human agent to intervene and manually implement alternative solutions that may better fit the situation at hand.

To summarize, the point is that, with increased levels of automation, it becomes even more important to show more information (i.e., opening the black box), not less. The goal here is to ensure that cooperating agents have a common ground that reflects the deep structure of the control problem in a productive way. Finding the common ground that underlies both the interface and the automation should thereby be

Fig. 6: The principle of relative navigation made explicit on a pilot’s navigation display. Here, the velocity obstacles of two intruder aircraft are rendered in the ownship maneuvering envelope, enabling a pilot to detect potential conflicts and observe the solution space to resolve conflicts in both speed and heading. (Figure taken from [74])
guided by a formative model of the work or problem domain, which is the core of the EID approach. In this context, Rasmussen’s abstraction hierarchy becomes a key framework for representing the results of work analysis [71]. Although the abstraction hierarchy in itself is not a ‘cookbook’ for displays and automation, it is a powerful critical-thinking tool that helps designers to structure functional relationships and constraints within the work domain in a way that can inform decisions about both displays and automation [78]. Thus, EID design is motivated by the hypothesis that many of the problems associated with ‘situation awareness’ and ‘trust in automation’ (e.g., [79]) result from the failure of interface designers to create common ground so that the operator can better understand the rationale guiding the behavior of automated team mates.

VII. THE FUTURE OF EID

The introduction of the EID perspective by Rasmussen and Vicente [9] has contributed to a paradigm shift in how the interface design problem is formulated. This shift was motivated by the observation that information technologies were changing the nature of work. On the one hand, the technologies were driving increased complexity (i.e., requisite variety), as people tried to take advantage of increased access to information and computational power to improve capabilities, efficiency, and resilience (e.g., as recognized by Hollnagel and Woods [5], Perrow [64], and others). In the aviation domain, for example, this is reflected in initiatives to explore new concepts of airspace management [25], [26] that take advantage of improved sensing (e.g., GPS), communication (e.g., digital data links) and display technologies (e.g., sense and avoid cockpit displays). On the other hand, the same technologies also offered increasingly flexible ways to integrate, organize, and represent the information to human operators (e.g., configurational graphics on a glass cockpit). As a result of these changes, the focus of interface design has shifted from designing to facilitate access to data (i.e., a problem of input or communication), to designing to improve problem solving and decision-making (i.e., a problem of supporting productive thinking). This shift in focus has also been marked by other new constructs, outside of EID, such as ‘joint cognitive systems’ [6], ‘multi-level flow modeling’ [80], and ‘direct manipulation interfaces’ [81].

The trends that motivated the introduction of EID and the related constructs are continuing and accelerating; work will become increasingly complex and the need for and opportunities to develop innovative representations to support productive thinking will grow. One consequence of this is that the need for a label like “EID” may fade, because both the demand and opportunities for innovation at the interface will become increasingly obvious. Thus, it will no longer be necessary to mark the distinction between the classical approaches that focused exclusively on data acquisition, and new approaches that emphasize productive thinking. Once one recognizes the demand and opportunity for the interface to support productive thinking, then there is little doubt that the interface must correspond with the problem domain being represented (requiring some type of ‘work domain’ analysis) and must do so in a way that shapes the operator’s thinking in productive ways (requiring some model of human cognition). Thus, as the old paradigm may fade, we expect that the “ecological” adjective will become redundant, and designers will measure display innovations relative to their capacity to support productive ways of thinking.

We have no doubt that the EID approach marks a paradigm shift (that was already in progress) with respect to how the problem of interface design is approached. Today the majority of designers recognize that interface design is no longer exclusively about access to data. So, the crucial question with respect to the EID initiative is whether the specific formalisms associated with EID (i.e., the abstraction hierarchy for work domain analysis and the SRK model of human cognition) will continue to inspire design innovation in the future. Our bet is the answer will be ‘yes.’ The abstraction hierarchy reflects basic insights into the nature of functional interactions, and provides a particularly useful way to partition the functional complexity of work domains in order to construct representations that capture the ‘deep structure’ (i.e., the semantics of work domain). This can thus be a basis for organizing (or chunking) information in both internal and external representations. Similarly, the SRK model provides unique insights into the nature of the coupling between cognitive agents and control problems. That is, the skill-based level recognizes the power of direct manipulation and perception through analog representations of continuous variables associated with control processes with small time constants. The rule-based level recognizes that as the effective time constants increase, stable control will often require more ballistic, pre-packaged responses to triggers in the work domain. Finally, due to the complexity of the domain, operators will often need to complete the design through more formal (knowledge-based) reasoning to diagnose problems or to discover and realize opportunities that were not anticipated in the original design of the system. All three types of coupling need to be considered and supported in the design of interface representations. It would, however, be foolish to claim that these tools are the only path to design innovation. It is therefore important that these formalisms not be dogmatized into procrustean beds. Rather, we expect that these tools will be adapted and refined to meet the special needs of particular work domains and design opportunities (just as varieties of hammers have evolved to fit the demands of different jobs).

In sum, we are confident that the direction set by the EID approach will increasingly be taken for granted as a productive approach to interface design. The goal for this review of the EID approach has not been to defend EID as a ‘brand’, but rather to endorse the value of the formalisms (the abstraction hierarchy and the SRK model) as useful tools for those seeking ways to support productive thinking through representation design. We predict that such formalisms will be increasingly utilized and refined to meet the demands and opportunities of future work domains. As a result, the adjective ‘ecological’ might become increasingly redundant, yet the focus on supporting productive thinking will become increasingly important as a paradigm for engineering representations. The success of EID will be complete when people simply recognize that ‘good’ interfaces are ones that reflect the ‘deep structure’
of work in order to support productive thinking.

REFERENCES


