2014 IEEE International Conference on Systems, Man, and Cybernetics
October 5-8, 2014, San Diego, CA, USA

Revisiting three ecological interface design experiments to investigate performance and control stability effects under normal conditions

Chelsea Carrasco, Greg A. Jamieson, Olivier St-Cyr
Cognitive Engineering Laboratory
Department of Mechanical and Industrial Engineering
University of Toronto, Toronto, Ontario, Canada
jamieson@mie.utoronto.ca

***Abstract*— This paper seeks to: (1) provide statistical evidence regarding operator task performance and control stability during the learning phase of two archival experiments in which participants in EID and non-EID interface groups were balanced by cognitive style; and (2) to compare task performance and control stability from these two experiments and a third archival experiment differentiated by the presence or absence of early faults and sensor noise. Participants in the EID condition of the first two studies 1) achieved target goals significantly faster, and 2) exhibited more stable control than those in the non-EID condition. When considered in context of the third study, the results again showed participants in the EID condition outperformed those in the non-EID condition. These results stand in contrast to previously reported findings, wherein no task performance differences were observed between these interface conditions.**

I. INTRODUCTION

A. Ecological Interface Design

Ecological Interface Design (EID [1]) is a framework for creating human-machine interfaces for complex systems. The framework employs an analysis of the physical and functional constraints of a work domain and a mapping of those constraints on to interface forms. Empirical studies have shown consistent benefits for EID interfaces in terms of improved detection, diagnosis and compensation for disturbances (see [2] for an interim review). Researchers have argued that providing an explicit depiction of work constraints in the interface aids these problem-solving activities. These same constraints are relevant under normal operating conditions as well. However, there is little evidence that EID interfaces support control task performance under normal operating conditions. Contrary to this trend, Jamieson [3] reported task performance benefits for EID interfaces under normal operating conditions. He suggested that the normal conditions in earlier studies might have lacked sufficient domain and/or task complexity to elicit the problem solving behavior that EID is most effective at supporting.

B. Cognitive Style

Torenvliet et al. [4] concluded that the strongest and most consistent predictor of operator task performance across five seminal EID studies was the cognitive style of the operator. *Cognitive style* refers to individual differences in learning strategies and information processing. Under this construct, individuals are categorized according to their predisposition to a *holist*, *serialist, or versatile* cognitive approach. Holistic thinkers will usually try to understand the overall principles of a task, and will test multiple hypotheses in parallel. Conversely, serialists test hypotheses sequentially [5]. Versatile thinkers are approximately equally likely to employ holist or serialist methods. Cognitive style can be assessed via the Spy Ring History Test [6]. The test yields scores in each of three dimensions: Holist, Serialist, and Neutral.

Torenvliet et al. [4] showed that the interaction of high Holist score and assignment to the EID interface predicted faster trial completion (i.e., better control task performance), whereas a high Serialist score and assignment to the non-EID interface predicted slower trial completion. For this reason, several subsequent EID studies have controlled for cognitive style as an individual difference.

C. Benefits of EID Interfaces under Normal Conditions

In this paper we consider three EID experiments that employ cognitive style as an experimental control. Table I provides a brief summary of each experiment.

TABLE I. EID EXPERIMENTS

Abbr. Name	Year	Author	Trials Used	Description
H&V1 [7]	2002	Hajdukiewicz & Vicente	Blocks 1-3 (of 4)	Between subjects experiment to test the impact of EID on adaptation to novelty and change using the DURESS II P and P+F interfaces.
H&V2 [8]	2004	Hajdukiewicz & Vicente	Blocks 1-3 (of 4)	
S&V [9]	2004	St-Cyr & Vicente	Blocks 1-3 (of 4)	

Henceforth, we will refer to these papers by their abbreviated names and reference numbers.

Participants in these experiments were assigned to either an EID or a non-EID interface condition such that pairs of participants with similar cognitive style were matched across interface levels. The non-EID interface, (called the P interface, see Fig. 1) displays primarily physical information about the work domain. In contrast, the EID interface (called the P+F interface, see Fig. 2) displays both physical and functional information about the work domain (in a cognitively relevant

978-1-4799-3840-7/14/$31.00 ©2014 IEEE

manner) by means of configural displays. Thus, it contains high-level emergent features based on low-level sensor data [9]. Working with the DURESS II thermal-hydraulic process control microworld, participants in all three studies completed three blocks of 20 trials in a learning phase prior to a fourth and final block introducing perturbations to the process [10].

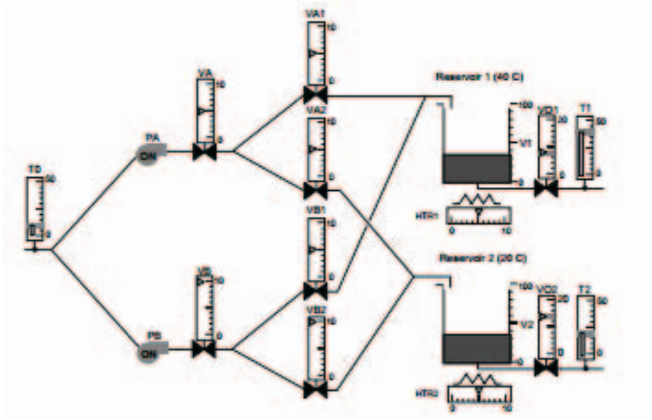

Fig. 1. P Interface on DURESS II [9]

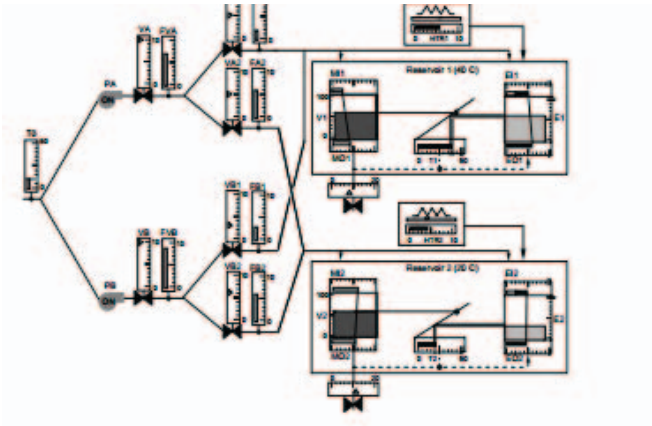

Fig. 2. P+F Interface on DURESS II [9]

H&V1 [7] reported both task performance and control stability advantages for the EID interface condition in the third block of normal trials (in the first of two experiments; the second of which is not pertinent). However, in keeping with their research objective, these results were only used as a benchmark for the perturbation trials.

H&V2 [8] reported only task performance results for the third block of normal trials. However, visual inspection of their Fig. 7 shows that the 95% confidence intervals barely overlap, suggesting that the difference between the interface groups approached significance. Despite this occurrence, they reported no statistically significant differences between the interface groups. Once again, however, task performance was not critical to the research objective and the finding was not discussed.

In another experiment, St-Cyr et al. [11] (which is not included in Table 1) reported improved operator task performance and control stability for the EID interface group across the entire learning phase of a third DURESS study. However, the non-EID interface employed by St-Cyr et al. [11] contained more information than the non-EID interface employed in H&V1 [7] and H&V2 [8]. Thus, while the findings are consistent, they are not easily compared.

Fortunately, another EID study conducted by St-Cyr and Vicente, S&V [9], affords a comparison of operator task performance and control stability across the learning phase of a DURESS II study employing a similar method and the same EID and non-EID interfaces as H&V1 and H&V2 [7, 8]. Once again, however, S&V [9] treated the learning phase results as a benchmark for the sensor noise perturbations introduced in their Block 4, and while normal trial data results were reported, they were neither focused upon nor discussed.

D. Objectives

In summary, DURESS II microworld studies wherein interface groups were not balanced by cognitive style have generally failed to demonstrate task performance or control stability benefits of EID interfaces under normal (i.e., non-disturbance) trials. Thus far, we have reviewed three DURESS II experiments wherein interface groups were balanced according to cognitive style. In H&V1 and H&V2 [7,8], participants using an EID interface showed some task performance and control stability advantages under normal operating conditions. S&V [9] showed a trend toward a task performance advantage for the EID interface group, but only for a single block of trials.

There are three key limitations to these findings. First, neither H&V1 [7], nor St-Cyr et al. [11] reported statistical results to substantiate the findings from the first two blocks of H&V1 [7]. Second, H&V2 [8] presented no results for control stability in their second study. Third, the non-EID interface conditions in H&V1 and H&V2 [7,8] are not the same as the non-EID interface in St-Cyr et al. [11], complicating any comparisons between the experiments.

This paper seeks to address those limitations. First, it provides the statistical evidence to substantiate the claim of better operator performance on the EID interface across the three-block learning phase observed (but not reported) by H&V1 or H&V2 [7,8]. Second, it compares the task performance and control stability findings from H&V 1 and 2 [7,8], and S&V [9]. These contributions yield new and important insight into the benefits of EID interfaces under normal operating conditions.

II. METHOD

The first part of this investigation analyzed the learning phase data (i.e. Blocks 1-3) from two experiments [7, 8] with 16 participants each. The second part consisted of selecting 10 (of 16) pairs of participants from H&V1 [7] and H&V2 [8] to compare to the 10 pairs of participants in S&V [9]. A brief overview of these experiments follows.

All three experiments were conducted using the DURESS II microworld, a representative thermal-hydraulic process simulation [9, 12]. In each experiment, participants were assigned to one of two interface groups: the non-EID interface (see Fig. 1) or the EID interface (see Fig. 2). Eight pairs of participants were assigned to each interface in the two experiments by H&V [7, 8], and ten pairs in S&V [9]. All participants were engineering students selected based on their willingness to participate, their degree of relevant formal training in physics (each participant had completed at least two university level courses), and their Spy Ring Test scores.

The task in each trial was to bring an idle process to steady state by satisfying four goal conditions for five consecutive minutes. These included temperature and flow rate constraints on water exiting a heated tank.

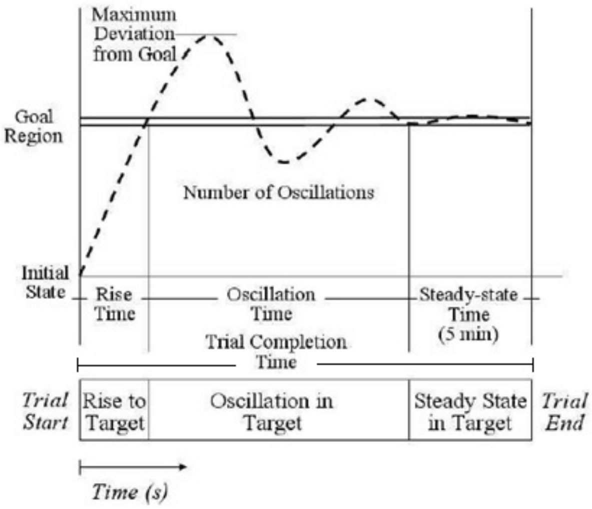

Fig. 3. Dependence Measures [9]

Task performance is assessed in terms of Trial Completion Time (TCT). In addition, Rise Time (RT), Oscillation Time (OT), Number of Oscillations (NO), and normalized Maximum Deviation (MD) were measured to assess stability of control (see Fig. 3 and Table II).

TABLE II. DEPENDENT MEASURES

Construct	Measure	Definition
Performance	***TCT***	Time (in seconds) it took for participants to reach the commanded steady state condition (control target) starting from a shutdown state.
Control Stability	***RT***	Elapsed time from the start of trial until all goal variables have reached the target range.
	NO	Number of times the goal variables crossed above and below the target regions.
	OT	Time (in seconds) to stabilize the goal variables after they reached the target regions and before steady state was achieved.
	MD	Maximum value by which the goal variables exceeded the target regions.

There are three methodological differences between the two Hajdukiewicz studies and the St-Cyr study: First, H&V1 and 2 [7, 8], included no sensor noise in the learning phase, whereas S&V [9] included industry average sensor noise. Second, H&V1 and 2 [7, 8] used a 0.1s (10 Hz) display refresh rate, whereas S&V [9] used a refresh rate of 2s (0.5 Hz). Third, H&V1 [7] incorporated faults in trials 2 and 4 of their Block 1 whereas S&V [9] did not include any fault trials in the learning phase.

A. *Statistical Analysis*

Part 1. For each of the H&V1 and 2 [7, 8] learning phase trial blocks, we calculated a 95% Confidence Interval (CI) of the mean amongst all 16 participants for each dependent variable, with interface as a between-subjects factor (P vs. P+F).

Part 2. We made a statistical comparison of each measure collected from St-Cyr & Vicente [9] and Hajdukiewicz and Vicente [7], [8] using the Type I General Linear Model (GLM) ANOVA with study and interface type as predictors. For this ANOVA, ten (of 16) pairs of participants were selected from H&V1 and 2 [7, 8] to provide equivalent statistical power to the ten pairs of participants in S&V [9]. Additionally, only trials 5-20 were considered in Block 1, since H&V1 [7] included fault conditions in Trials 2 and 4 of Block 1. Post-hoc tests were performed if the GLM showed significant results for either Interface or Study for any of the measures.

To choose ten pairs of participants from H&V1 [7] and H&V2 [8], we first determined the Spy Ring Test distance score between each of the original 16 pairs. The distance score of each pair was calculated by summing the squares of the differences between the Holist, Serialist and Neutral scores of each participant in the pair, and then taking the square root of that value. We then identified the 10 pairs from [7] and [8] whose score most closely matched that of a pair in S&V [9]. For each pair in S&V [9], the algorithm identified which pair from [7] and [8] was most similar. When two or more pairs from [9] shared the same closest match among the pairs in [7] and [8], then the set of pairs that was most similar were grouped. This process was repeated until all pairs from S&V [9] were matched. This ensured that the two studies provided equivalent statistical power, and that the quality of the matches between the two groups was approximately equivalent. See Table III for the results of the pair selection.

Although S&V [9] pairs 4 and 6 appear to reflect relatively poor matches in comparison to their peers, the algorithm used to select matches creates an overall close match selection at the perceived expense of some matches. Moreover, of the S&V [9] pairs matched, 4 and 6 were still considered relatively close matches; it was not unusual to encounter pairs with differences exceeding 60 units.

TABLE III. PARTICIPANT MATCH SELECTION

St-Cyr and Vicente [9*]	Score	Hajdukiewicz and Vicente [7, 8]	Score
Pair 1	16.67	Pair 1, Experiment 2	16.58
Pair 2	4.00	Pair 7, Experiment 1	4.00
Pair 3	1.41	Pair 2, Experiment 1	1.00
Pair 4	3.61	Pair 5, Experiment 2	10.10
Pair 5	9.64	Pair 8, Experiment 1	9.95
Pair 6	1.41	Pair 7, Experiment 2	10.34
Pair 7	9.06	Pair 2, Experiment 2	9.90
Pair 8	8.60	Pair 1, Experiment 1	7.55
Pair 9	10.86	Pair 5, Experiment 1	10.77
Pair 10	5.92	Pair 3, Experiment 2	5.00
Average	**7.12**	**Average**	**8.52**

III. RESULTS

A. Part 1

95% confidence intervals of the mean for TCT, OT, NO, and RT measures are shown in Figures 4-7, respectively. Results from these measures are generally consistent with the findings reported in St-Cyr et al. (see Footnote 3 of [11]). None of the measures yielded normally distributed results, nor did the Box-Cox transformations result in normally distributed data, therefore ANOVA results are not reported.

Participants in the P+F condition consistently and unambiguously completed trials faster in each study, for each block of trials (see Fig. 4). The magnitude of the difference between TCT performance on the non-EID interface and the EID interface is consistently between 100-200 seconds better on the EID interface, on a task that at worst took under 1100 seconds to complete, and at best under 500. OT and NO also improved consistently and unambiguously for participants in the P+F condition.

RT was consistently lower for P+F participants in H&V1 [7], however it was consistently higher for P+F participants in H&V2 [8] and S&V [9] as shown in Fig. 7.

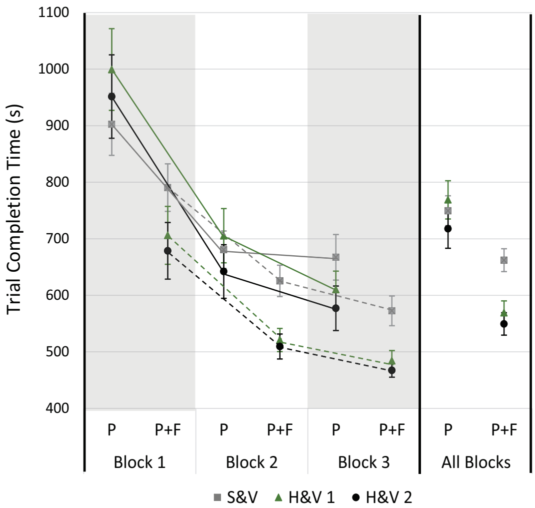

Fig. 4. Interval Plot of Mean TCT (s) – 95% CI

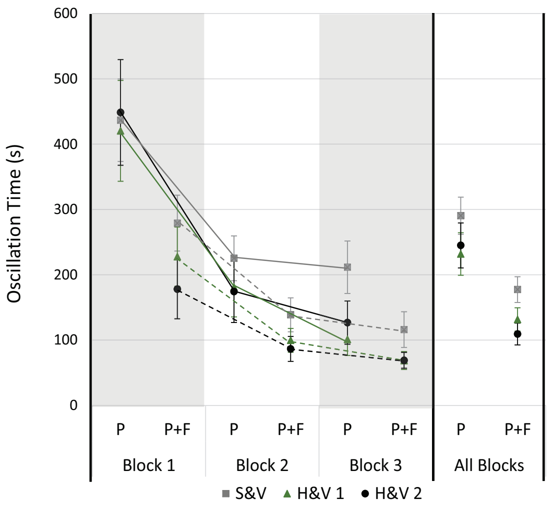

Fig. 5. Interval Plot of Mean OT (s) – 95% CI

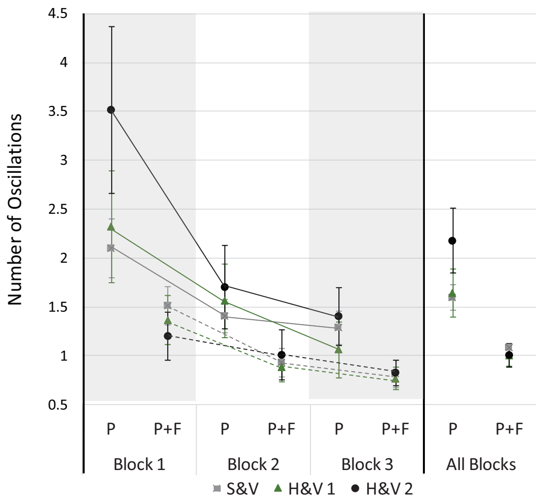

Fig. 6. Interval Plot of Mean NO – 95% CI

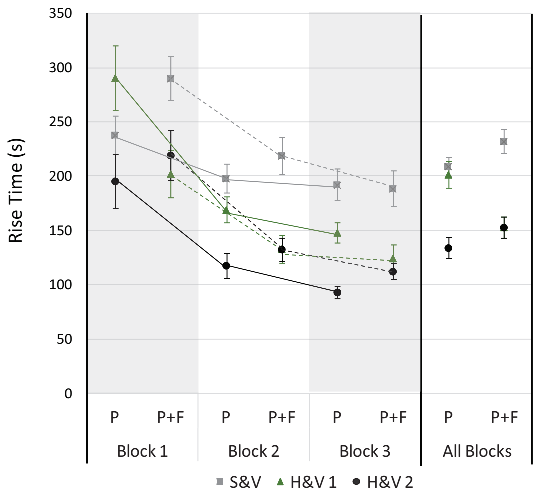

Fig. 7. Interval Plot of Mean RT – 95% CI

B. Part 2

The results of the General Linear Model on TCT yielded a main effect of Interface, $F(1, 2215) = 124.02$, $p < 0.001$, such that the average TCT was significantly faster for participants in the P+F condition (M = 567.8) than for those in the P interface condition (M = 693.3). The main effect of Study was also significant, $F(1, 2215) = 75.85$, $p < 0.001$, such that the average TCT was significantly faster for the selected participants from H&V [7, 8] (M = 581.5) than for participants in S&V [9] (M = 679.6).

The interaction effect between Interface and Study was also significant, $F(1, 2215) = 17.37$, $p < 0.001$. The difference in TCT between groups of P and P+F participants in the H&V studies [7, 8] was greater than that observed in S&V [9]. In other words, H&V participants appeared to benefit more from the P+F interface [7, 8]. See Table IV for a summary of all GLM p-values).

TABLE IV. GENERAL LINEAR MODEL RESULTS – ALL BLOCKS

	Interface		Study		Interface* Study	
	p-value	*F-stat*	*p-value*	*F-stat*	*p-value*	*F-stat*
TCT $F_{1,2215}$	<0.001	124.02	<0.001	75.85	<0.001	17.37
RT $F_{1,2213}$	0.007	7.30	<0.001	398.29	0.014	6.10
OT $F_{1,2212}$	<0.001	98.53	<0.001	50.25	0.337	0.92
NO $F_{1,2212}$	<0.001	85.54	0.315	1.01	<0.001	12.75
MD $F_{1,2211}$	0.067	3.36	0.026	4.97	<0.001	16.96

Post-hoc comparisons using the Tukey HSD test confirmed the above findings and further indicated that the mean TCT for the P+F condition in S&V [9] was not significantly different from the P condition in the Hajdukiewicz and Vicente studies [7, 8]. However, all other treatment combinations were significantly different; the Tukey Test results denote this by assigning different alphabetical groupings to statistically different treatment combinations (See Table V).

TABLE V. TCT TUKEY TESTS - (95% CONFIDENCE)

Interface	Study	Samples	Mean	Grouping
P+F	-	1109	567.8	A
P	-	1108	693.3	B
-	Hajdukiewicz	1108	581.5	A
-	St-Cyr	1109	679.6	B
P+F	Hajdukiewicz	553	495.3	A
P	Hajdukiewicz	555	667.7	B
P+F	St-Cyr	556	640.4	B
P	St-Cyr	553	718.8	C

The Tukey Test results for RT, OT, NO and MD are shown in Tables VI-IX, respectively.

TABLE VI. RT TUKEY TESTS – 95% CONFIDENCE

Interface	Study	N	Mean	Grouping
P+F	-	1107	178.7	A
P	-	1107	167.2	B
-	St-Cyr	1109	215.6	A
-	Hajdukiewicz	1105	130.3	B
P+F	St-Cyr	556	226.7	A
P	St-Cyr	553	204.6	B
P	Hajdukiewicz	554	130.7	C
P+F	Hajdukiewicz	551	129.8	C

TABLE VII. OT TUKEY TESTS – 95% CONFIDENCE

Interface	Study	Samples	Mean	Grouping
P	-	1107	226.76	A
P+F	-	1107	114.07	B
-	St-Cyr	1109	210.65	A
-	Hajdukiewicz	1105	130.18	B
P	St-Cyr	553	261.55	A
P	Hajdukiewicz	554	191.97	B
P+F	St-Cyr	556	159.76	B
P+F	Hajdukiewicz	551	68.38	C

TABLE VIII. NO TUKEY TEST RESULTS – 95% CONFIDENCE

Interface	Study	Samples	Mean	Grouping
P	-	1107	1.7577	A
P+F	-	1107	0.8947	B
-	Hajdukiewicz	1105	1.373	A
-	St-Cyr	1109	1.2793	B
P	Hajdukiewicz	554	1.9711	A
P	St-Cyr	553	1.5443	B
P+F	St-Cyr	556	1.0144	C
P+F	Hajdukiewicz	551	0.775	C

TABLE IX. MD TUKEY TESTS – 95% CONFIDENCE

Interface	Study	Samples	Mean	Grouping
P+F	-	1106	1.3882	A
P	-	1107	0.986	A
-	St-Cyr	1109	0.9425	A
-	Hajdukiewicz	1104	1.4317	B
P	Hajdukiewicz	554	0.7787	B
P+F	Hajdukiewicz	550	2.0847	A
P	St-Cyr	553	1.1934	B
P+F	St-Cyr	556	0.6917	B

IV. DISCUSSION

The two objectives of this study were:

1. To provide statistical evidence regarding operator task performance and control stability during the learning phases of H&V1 [7] and H&V2 [8]; two studies in which participants in EID and non-EID interface groups were balanced by cognitive style, and
2. To compare task performance and control stability between the two H&V experiments [7, 8], and S&V [9].

Two important findings emerge. First, across the learning phases of H&V1 [7] and H&V2 [8], participants using an EID interface performed significantly better on the control task than participants using a non-EID interface. Second, participants using the EID interface generally exhibited more stable control.

The first finding echoes those reported in St-Cyr et al. [11]. This may be attributed to both studies having applied the minimum distance calculator to the Spy Ring test scores to match participant cognitive style across interface conditions. Recall that Torenvliet et al. [4] demonstrated that the interaction of high Holist score and assignment to the EID interface predicted faster trial completion; whereas the interaction of high Serialist score and assignment to the non-EID interface predicted slower trial completion. It is possible that in prior DURESS studies where no main effects of interface were observed under normal operating condition, high Holist score individuals were under-represented in EID interface conditions, and/or that high Serialist individuals were under-represented in the non-EID interface conditions. We are currently investigating this possibility to determine whether such an imbalance may have biased the results.

The second important finding arises from the marked magnitude difference in task performance and control stability measures between S&V [9] and H&V [7, 8]. On all measures, participants in H&V1 [7] and H&V2 [8] outperformed those in S&V [9]. These differences may be attributed to one of the three methodological differences between the two studies (sensor noise, refresh rate, or presence of fault trials), as outlined in the methodology section. The first two differences suggest that participants in S&V [9] simply encountered a more difficult control task. With respect to the third difference, the introduction of faults in early trials by H&V [7, 8] may have accelerated the development of participants' "deep knowledge" [13].

V. CONCLUSION

Complex systems operate over a range of normal and abnormal conditions. However, the human factors literature in general – and the EID literature in particular – gives disproportionate emphasis to performance evaluation under abnormal or emergency conditions. The findings discussed here take advantage of historical EID studies to investigate whether EID interfaces support control task performance and stability under normal operating conditions. This affords a more complete assessment of whether EID interfaces holistically support operators of complex systems.

The results support the conclusion that EID interfaces support improved control task performance and greater control stability under normal operating conditions – a conclusion that is not well supported in previous EID literature. Coupled with the more widely reported advantages of EID under fault conditions, these results might positively influence the adoption of EID interfaces in industry.

Finally, whereas the individual cognitive style of the operators was controlled for in the studies revisited by this paper, we have little understanding of the mechanism by which cognitive style influences control performance. Moreover, we have no evidence of the robustness of the cognitive style effect across other experimental settings, interventions, or operator populations. Future studies should seek to discover this mechanism and explore the breadth of its implications for human performance in complex systems. If the effects persist, cognitive style might be considered a criterion for operator selection when EID interfaces are employed.

VI. REFERENCES

[1] K.J. Vicente and J. Rasmussen, "Ecological interface design: theoretical foundations," *IEEE Trans.Syst.Man Cybern.,* vol. 22, pp. 589–606, 1992.

[2] K.J. Vicente, "Ecological interface design: Progress and challenges," *Human Factors*, vol. 44, pp. 62–78, 2002.

[3] G.A. Jamieson, "Ecological interface design for petrochemical process control: an empirical assessment," *IEEE Trans.Syst. Man Cybern. A., Syst. Humans*, vol. 37, pp. 906-920, 2007.

[4] G.L. Torenvliet, G.A. Jamieson, and K.J. Vicente, "Making the most of ecological interface design: The role of individual differences," *Applied Ergonomics*, vol. 31, pp. 395-408, 2000.

[5] G. Pask, "Styles and strategies of learning," *British Journal of Educational Psychology,* vol. 46, pp. 128-148, 1976.

[6] G. Pask and B.C. Scott, "Learning styles and individual competence," *International Journal of Man-Machine Studies*, vol. 4, pp. 217–253, 1972.

[7] J.R. Hajdukiewicz and K.J. Vicente, "Designing for adaptation to novelty and change: functional information, emergent feature graphics, and higher-level control," *Human Factors,* vol. 44, pp. 592-610, 2002.

[8] J.R. Hajdukiewicz and K.J. Vicente, "What does computer-mediated control of a thermal-hydraulic system have to do with moving your jaw to speak? Evidence for synergies in process control," *Ecological Psychology*, vol. 16, pp. 255-285, 2004.

[9] O. St-Cyr and K.J. Vicente, "Sensor noise and ecological interface design: effects on operators' control performance," In *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting, HFES 2004, Santa Monica, CA, USA*, pp. 538-42, 2004

[10] W. Pawlak and K.J. Vicente, "Inducing effective operator control through ecological interface design," *International Journal of Human-Computer Studies*, vol. 44, pp. 653-688, 1996.

[11] O. St-Cyr, G.A. Jamieson, and K.J. Vicente, "Ecological interface design and sensor noise," *International Journal of Human-Computer Studies*, vol. 71, pp. 1056-1068, 2013.

[12] K.J. Vicente and J. Rasmussen, "The ecology of human-machine systems II: Mediating "direct perception" in complex work domains," *Ecological Psychology*, vol. 2, pp. 207-250, 1990.

[13] K. Christoffersen, C.N. Hunter, and K.J. Vicente, "A longitudinal study of the impact of ecological interface design on deep knowledge," *International Journal of Human Computer Studies,* vol. 48, pp. 729-762, 1998.