

# Levels of automation in human factors models for automation design: Why we might consider throwing the baby out with the bathwater

**Greg A. Jamieson  
Gyrd Skraaning Jr**

## Abstract

This paper responds to Kaber's reflections on the empirical grounding and design utility of the Levels of Automation (LOA) framework. We discuss the suitability of the existing human performance data for supporting design decisions in complex work environments. We question why human factors design guidance seems wedded to a model of questionable predictive value. We challenge the belief that LOA frameworks offer useful input to the design and operation of highly automated systems. Finally, we seek to expand the design space for human-automation interaction beyond the familiar human factors constructs. Taken together, our positions paint LOA frameworks as abstractions suffering a crisis of confidence that Kaber's remedies cannot restore.

## Differing perspectives on common ground

Professor David B. Kaber's position paper invites a welcome exchange of ideas about a central construct in human factors engineering: Levels of Automation (LOA). We applaud JCEDM for airing a range of responses to his views in the same issue. This approach offers a promising contrast to slow motion, ping pong exchanges between journals with competing perspectives on human factors science.

The authors collaborate on human-automation interaction research as both empiricists [Burns et al., 2008; Lau et al., 2008a, 2008b; Lau, Jamieson & Skraaning, 2014, 2016a, 2016b] and designers [Hurlen, Skraaning, Myers, Jamieson & Carlson, 2015; Jamieson, Hurlen & Skraaning, 2014; Skraaning, Hurlen, LeDarz & Jamieson, 2016]. We adopt an inductive approach to research, building knowledge through prototyping and experimentation in complex work environments. We seek to support designers by aligning that knowledge with the richness of the design problem, and not primarily through models of questionable predictive value. This perspective has evolved through twenty years of realistic simulator studies of human-automation interaction on complex process control tasks; where human factors models have shown little utility in predicting performance outcomes (Skraaning & Jamieson, in preparation). It is from this perspective that we respond to Kaber's comments on the empirical basis for, and design relevance of, LOA models.

We concur with much of Kaber's critical consideration of the empirical evidence for LOA predictions. We agree that human factors research should move away from constrained

artificial situations and toward the problems faced by designers of real operating environments. As he proposes, existing models of types and levels of automation based on human information processing may not be descriptive of operator behavior in complex systems, yielding a frustrating discrepancy between the models' predictions and real-life observations. As Kaber points out, human-automation interaction outcomes are likely a function of more than human information processing. We heartily agree to his opening up of the model space to encompass teamwork, trust, and other metaphors.

## On the need for, and utility of, human performance data in support of LOA-based predictions

Fitts et al. (1951) recognized a need for a research program to create the human performance data to effectively compare human and machine execution of tasks. Yet nearly 35 years after they introduced MABA-MABA and levels of automation, Price (1985) cited a "general weakness of applicability" (p.35) of function allocation methods. These methods: presumed that human performance data would exist from which the performance of humans could be predicted... [They] would depend on the availability of large quantities of quantified data on human performance, data that could be calibrated to the specific conditions of a new design. *Such data do not now exist, and they probably never will.* (Price, 1985, pp. 35-36, emphasis added.)

Today, more than 65 years since the genesis of the Fitts List, the theory of levels of automation, and the recognition of the need for large quantities of human performance data, Price's prediction has proved prescient. LOA models depend on data that we largely do not have. And like him, although we share the appetite for more data, we cannot envision the human factors community mustering the resources needed to supply it.

We interpret Kaber's call for finer grained LOA models as an effort to more precisely specify the data needed to validate new deductive models for increasingly complex systems. Like Rouse (1988), we have reservations about the practicality of using empirical data to deductively predict task performance for any reasonably complex system. We also share Perrow's (1984) perspective that simple and complex technological systems differ in kind as opposed to degree. Such systems are essentially unpredictable; their behaviors emergent. In such settings, we argue, an inductive research approach is preferred to build knowledge and extract design principles.

## On the strength of evidence in support of LOA-based predictions in complex work settings

Kaber expresses concern about the inconsistent response trends in individual LOA experiments. Onnasch et al. (2014) applied reasonable selection criteria to a much larger body of research to identify studies for inclusion in their meta-analysis. Those 18 studies provide a proxy for a discussion about the strength of the empirical support for LOA models. However, we argue that the composition of the studies themselves fails to support the application of the results to complex work systems.

Most of the studies included in the meta-study were conducted in laboratory multi-task contexts. According to Onnasch et al. (2014, p. 485), only four of the 18 studies included non-student participants; one of which employed military personnel not engaged in tasks pertaining to their expertise (i.e., Calhoun et al. 2009). Similarly, Cummings and Mitchell (2007) recruited active duty military personnel with overlapping subject-matter expertise with respect to an anticipated future-world experimental task (Cummings, personal communication, March 31, 2017). The participants in Metzger and Parasuraman (2005) and Sarter and Schroeder (2001) were en route controllers and commercial aircraft pilots, respectively.

Table 1 extracts from Onnasch et al. (2014, p. 481) the studies that employed non-student participants and introduces results from an experiment that we are currently preparing for dissemination (Skraaning & Jamieson, in preparation). We have also added a column noting the experimental characteristics of each study, including the participants, the simulation environment, and the experimental tasks or scenarios. Table values are Kendall's Rank Correlation Coefficients, referred to as Kendal's Tau; a non-parametric measure of correlation. DOAs in each study were converted to sequential rank values and compared to dependent meta-variable indicators distinguished in rank by statistical significance. For Skraaning & Jamieson (in preparation), we followed the ranking method as described in the meta-study (although, we have lingering questions about the validity of Onnasch et al.'s (2014) ranking method to support inferences about DOA effects on performance, workload or SA.)

Table 1. Selected (and supplemented) results from Onnasch et al. (2014) pertaining to findings from complex work environments.

Study	Experiment Characteristics	Routine Primary Task Perform.	Return-to-Manual Primary Task Perform.	Routine Secondary Task Perform.	Return-to-Manual Secondary Task Perform.	Subjective Workload	SA
Calhoun et al. (2009)	Military personnel (non-SME); Commercial simulator; complex scenarios	-0.816		0			0
Cummings & Mitchell (2007)	Active-duty military personnel; laboratory simulator; futuristic scenarios	0					0
Metzger & Parasuraman (2005)	En route controllers; medium fidelity task simulator; multi-task scenarios	0	0	0	0	0	0
Sarter &	Commercial aircraft	1					

Schroeder (2001)	pilots; Full-scope simulator; Complex scenarios						
Skraaning & Jamieson (in preparation)	Licensed NPP operators; Full-scope simulator; Complex scenarios	0				0	1 <sup>1</sup>

In stark contrast to the conclusions of Onnasch et al. (2014), Table 1 reveals little empirical evidence for a predictive model of LOA effects on task performance, situation awareness or workload for complex work settings. Ironically, removing these studies from the Onnasch et al. (2014) meta-study would presumably improve the predictive power of LOA characterizations for laboratory tasks that can be executed by students after a few hours of training (cf., Li et al, 2014).

Onnasch et al.'s (2014) findings do support the position that human factors engineers might use LOA predictions to make valid, reliable and useful predictions about automation design for simple work environments. However, their article offers no caution regarding limiting the application of the predictions to complex work settings given the current paucity of data.

## On the adherence to LOA-based predictions despite their poor predictive power

Although Kaber acknowledges the conflicting results from individual LOA experiments, he gives two reasons for standing behind predictive modeling of LOA as the general research strategy to support automation design. First, he argues that they are “handy”. But how handy can models of poor predictive value be? Do we have sufficient confidence in these models to encourage their use in the design of safety-critical applications?

By admitting that LOA-models developed and evaluated for seven decades are imprecise and unreliable (as Kaber does), it becomes hard to believe in the practical usefulness of the approach. Engineers will continue to make difficult HAI design decisions without the trustworthy technical basis promised by advocates of LOA models. We sympathize with the desire to offer consistent and universal human response trends that are truly useful to designers, but suspect that the LOA approach has overplayed its role in this regard.

Kaber, on the other hand, calls for a research program to establish fine-grained LOA models under the assumption that more sophisticated classification will lead to descriptive performance predictions and thereby greater utility to designers. We are skeptical of the assumption that more detail in these models will yield more accurate predictions than they have to date.

The second reason that Kaber remains committed to LOA look-up tables is that they are the best human factors researchers have to offer to designers and we should be reluctant to

---

<sup>1</sup> Linear contrast analysis showed a large experimental effect (partial eta-squared for four levels of procedure automation,  $\eta^2 = 0.30$ )

dispose of them without a suitable replacement at hand. In our view, this argument is false and impedes the search for better alternatives. It is an argument in favour of function allocation that has long been refuted by a minority in our community (e.g., Fuld, 2000):

...when a scientific discipline finds itself in a dead end, despite hard and diligent work, the dead end should probably not be attributed to lack of knowledge of facts, but to the use of faulty concepts which do not enable the discipline to order the facts properly. The failure of human factor engineering to advance in the area of allocation of function seems to be such a situation. (Jordan, 1963, pp.161-162.)

Kaber himself asserts that, "If actual performance observations on LOAs cannot be connected to theoretical descriptions, then such descriptions have little import in terms of systems design." Kaber, p.19. We agree with Kaber's appraisal of the quality of the empirical efforts to generate knowledge about human-automation interaction. We see no systematic flaw in the research methods employed to create the body of evidence. Rather, like Jordan (1963), we suspect that the poor predictive power of LOA-based predictions of human behavior in the presence of automation is based on faulty concepts. We do not see how greater refinement of these concepts will lead to more predictive models.

From our perspective, Kaber adopts a remarkably high threshold for rejecting the LOA concept. We should hesitate to throw this LOA baby out with the bathwater he says metaphorically. We concur with Kaber that LOA modeling is in its infancy: A 65-year old infancy. The LOA concept is the Benjamin Button of human factors research. After 65 years of imprecise and unreliable predictions, what more reason do we need to at least entertain the idea of throwing out this baby that fails to thrive?

## On the use and rejection of LOA taxonomies in system design and operation

Kaber points to the SAE's recent adoption of a taxonomy and definitions of driving automation systems (SAE, 2016) as evidence for their utility. However, other practitioner communities have expressed difficulty in applying the LOA concept. For example, The FAA's report on Operational Use of Flight Path Management Systems cites a limited utility in hierarchical LOA descriptions of flight deck automation.

The [working group] found that several operators started with a policy that used explicit definitions of levels of automation described as a simple hierarchy in a rigid and prescribed fashion. After gaining operational experience with training and operational use of these rigid definitions, several airlines concluded that such a description assumed a linear hierarchy that does not exist. The various features of the autoflight system (autopilot, flight director, autothrottle/autothrust, FMS, etc.), can be, and are, selected independently and in different combinations that do not lend themselves to simple hierarchical description. As a result of this experience, those operators revised their policies to allow the pilot to use the appropriate combination of automation features for the situation, *without rigidly defining them in terms of levels*. (Abbott, McKenney & Railsback, 2013, P. 55, emphasis added)

This example is noteworthy for two reasons. First, the critique emphasizes how the blend of automation capabilities in complex operating environments overwhelms simplistic

hierarchical descriptions. We have observed similar difficulties in the process industries as well, where the integration of individual automatic devices such as protections, controllers, scripts, routines and programs makes assignment of automation configurations to ordinal levels of a hierarchy a speculative endeavour. Instead of a single hierarchy, human factors researchers might appeal instead to multiple LOA hierarchies defined by stages of automation (Parasuraman, Sheridan & Wickens, 2000), system functions (autopilot, flight director...), or otherwise. Furthermore they might adopt the notion of adaptive selection of LOAs depending on the context of dynamic operational circumstances or operator state (Byrne & Parasuraman, 1996). In doing so we risk gravitating toward a descriptive theory of unique function allocation situations instead of providing general LOA predictions that are useful to designers.

Second, the FAA example highlights the value of *operational experience* with rigid LOA descriptions. In contrast, the SAE standard describes a simplified abstraction agreed to by stakeholders. In Kaber's words, the standard is a social construct. Looking beyond the six levels of automation reveals trade-offs in the taxonomy's negotiated formulation. The taxonomy excludes active safety systems and driver assistance systems such as automated emergency braking and lane keeping assistance, respectively (SAE, 2016, p.2). Thus, although the SAE's notion of the dynamic driving task includes longitudinal and lateral motion control, automation systems that intervene in these functions are excluded from the taxonomy. While this may be a useful simplifying assumption, a skeptic might wonder how effective such an abstraction can be in providing a framework for design specification and regulatory practice.

A second example of rejection of the LOA notion was provided by Dr. Clinton D. Chapman, Production Group Chief Software Architect at Schlumberger. He spoke in a session on Automation Case Studies at the September, 2014 Society of Petroleum Engineers Workshop on "Implementation of Drilling Systems Automation". Dr. Chapman reviewed both the Endsley and Kaber (1999) and Sheridan and Verplank (1978) LOA descriptions and recounted how the Drilling Systems Automation Technical Section had been "...reviewing this perspective over the previous 5 to 6 meetings...". He then asked if any of the workshop attendees was finding the concept useful. Not a single positive response was offered. The first author [Jamieson] was nonplussed. He rose to ask Dr. Chapman if he had understood correctly that the Technical Section was finding the LOA concept to be of no use in supporting the development and implementation of drilling automation. Chapman's reply was that the LOA concept was considered an abstraction that was not "fit for purpose" for drilling automation design.

It should come as no surprise that practitioners are split on the utility of the LOA concept. Like other human factors notions, LOA-based taxonomies offer an intuitive description and an apparently concise aid to cope with a wicked design problem. However, implementing the taxonomy can lead to the realization that the abstraction does not survive beyond the conceptual design phase. It remains to be seen whether the SAE taxonomy will enjoy greater success in surface transportation than reported in aviation and drilling.

## On the prospects for new thinking in HAI design

We agree with Kaber that HAI research asks some of the right design questions. But the LOA tradition prematurely closes off other questions. For example, it implies that the answer to, “How should automation interact with humans?” should be “by throttling the LOA (i.e., adjusting function allocation) and monitoring task performance, situation awareness and workload effects”, as if this was the only alternative. From our perspective, the teamwork metaphor (Christoffersen & Woods, 2002) and Dekker and Woods’ (2002) “abracadabra” accusation are responses to this unnecessary narrowing of the design space.

There are many more questions that LOA research does not address - questions that are discovered both in operational settings via automation “surprises” (Sarter & Woods, 1997) and through design experiences (Guerlain, Jamieson, Bullemer & Blair, 2002). Questions arising from an effort to develop concepts of operation for highly automated nuclear plants (Jamieson et al., 2014) include:

- Should we design for human interaction with many individual automated agents working at the component or sub-system level? Or through a single meta-agent?
- How should the designer allocate dialogue with machine agents amongst human agents? Should these dialogues be discretely held between one crew member and the agent, or held in the open? Should dialogue with automated agents be the role of one operator or shared amongst the crew?

We found no useful guidance on these questions in existing LOA research. Designers cannot wait for empiricists to discover the relevant questions through theory-driven research.

Of greater concern is that Kaber leaps from the assurance that we have the right questions to the assumption that we’re answering them in the right way. To the extent that he is willing to consider new approaches to HAI modeling, he assumes that starting over means returning to human information processing models (p. 22). Why make this assumption? After trying a single strategy for a long time with questionable success, should researchers fiddle with the same basic recipe? Or instead demonstrate some creativity in coming up with new ways of supporting designers of HAI in complex work environments?

At the same time let us also not succumb to nuevo-folk models of HAI that fixate on a single construct or mechanism. While constructs such as teamwork, trust and the like almost certainly serve as factors in predicting HAI, if taken in isolation they also threaten to close off degrees of design freedom. Engineers working on under-specified design problems must adopt a pragmatic approach. Like Sheridan’s (2000) bridge builder, they must start out with subjective intuitions before they can move on to normative/objective tools. Likewise, we human factors researchers should also be pragmatic and not subscribe to narrow models.

## Conclusion

In our view the LOA-paradigm has lost its momentum and is approaching a crisis. New HAI challenges are emerging along with continual technological development. It is not evident that automation is evolving in such a way that the LOA framework accurately describes the new design problems. We suspect that it is too late to save the concept through refinement or by hunting for stronger evidence of its predictive power. The failure of LOA theory thus far to account for human behaviour and experience in operational environments is as likely

attributable to faulty concepts as to insufficient facts (Jordan, 1963). If the evidence doesn't support the predictions of LOA models for safety- and production-critical work, we must be willing to at least consider rejecting the theory. We cannot persist in our unwillingness to be guided by accumulated evidence that forms a clear pattern over time just because we lack a contingency plan. And we cannot dismiss the critics for having nothing better (or even nothing at all) up their magician's sleeves.

## Acknowledgement

Professor Kaber shared many of the ideas expressed in his paper as a speaker in a Distinguished Seminar Series at the University of Toronto in December of 2016. He graciously agreed to Jamieson sharing an audio recording of his presentations with Skraaning. We extend our thanks to him for that collegial gesture; our comments are offered in the same spirit.

## References

- Abbott, K., McKenney, D., & Railsback, P. (2013). Operational use of flight path management systems. Final Report, Performance-based operations Aviation Rulemaking Committee/Commercial Aviation Safety Team, Flight Deck Automation Working Group.
- Burns, C. M., Skraaning Jr., G., Jamieson, G. A., Lau, N., Kwok, J., Welch, R. & Andresen, G., "Evaluation of Ecological Interface Design for Nuclear Process Control: Situation Awareness Effects," *Human Factors*, vol. 50, pp. 663-679, 2008.
- Byrne, E. A., & Parasuraman, R. (1996). Psychophysiology and adaptive automation. *Biological Psychology*, 42(3), 249-268.
- Calhoun, G. L., Draper, M. H., & Ruff, H. A. (2009, October). Effect of level of automation on unmanned aerial vehicle routing task. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 53, No. 4, pp. 197-201). Sage CA: Los Angeles, CA: SAGE Publications.
- Christoffersen, K., & Woods, D. D. (2002). How to make automated systems team players. In *Advances in human performance and cognitive engineering research* (pp. 1-12). Emerald Group Publishing Limited.
- Cummings, M. L., & Mitchell, P. J. (2007). Operator scheduling strategies in supervisory control of multiple UAVs. *Aerospace Science and Technology*, 11(4), 339-348.
- Dekker, S. W., & Woods, D. D. (2002). MABA-MABA or abracadabra? Progress on human-automation co-ordination. *Cognition, Technology & Work*, 4(4), 240-244.
- Endsley, M. R. & Kaber, D. B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, 42(3), 462-492.

Fitts, P. M., Ed. (1951). Human Engineering for an Effective Air Navigation and Traffic Control System (Chapter 3). Washington: National Research Council.

Fuld, R. B. (2000). The fiction of function allocation, revisited. *International Journal of Human-Computer Studies*, 52(2), 217-233.

Guerlain, S., Jamieson, G. A., Bullemer, P., & Blair, R. (2002). The MPC Elucidator: A case study in the design for human-automation interaction. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 32(1), 25-40.

Hurlen, L., Skraaning, G., Myers, W. P., Jamieson, G. A. & Carlson, H. (2015). "The Plant Panel: Feasibility Study of an Interactive Large Screen Concept for Process Monitoring and Operation." In *Proceedings of the 9th International Topical Meeting on Nuclear Plant Instrumentation, Control & Human-Machine Interface Technologies* (no pagination). La Grange Park, IL: American Nuclear Society.

Jamieson, G.A., Hurlen, L., & Skraaning Jr., G. (2014). *Highly Automated Plants: Perspectives, Methods and Prototypes (HWR-1128)*. OECD Halden Reactor Project.

Jordan, N. (1963). Allocation of functions between man and machines in automated systems. *Journal of applied psychology*, 47(3), 161.

Lau, N., Jamieson, G. A., & Skraaning Jr, G. (2014). Inter-rater reliability of query/probe-based techniques for measuring situation awareness. *Ergonomics*, 57(7), 959-972.

Lau, N., Jamieson, G. A., & Skraaning Jr, G. (2016a). "Empirical Evaluation of the Process Overview Measure for Assessing for Assessing Situation Awareness in Process Plants." *Ergonomics*, 59, 393-408.

Lau, N., Jamieson, G. A., & Skraaning Jr, G. (2016b). Situation awareness acquired from monitoring process plants—the Process Overview concept and measure. *Ergonomics*, 59, 976-988.

Lau, N., Veland, O., Kwok, J., Jamieson, G. A., Burns, C. M., Braseth, A. O., & Welch, R. (2008a), "Ecological Interface Design in the nuclear domain: An application to the secondary subsystems of a boiling water reactor plant simulator.," *IEEE Trans. Nuclear Science*, 55, 3579-3596.

Lau, N., Jamieson, G. A., Skraaning jr., G., & Burns, C. M. (2008b), "Ecological Interface Design in the nuclear domain: An empirical evaluation of ecological displays for the secondary subsystems of a boiling water reactor plant simulator.," *IEEE Trans. Nuclear Science*, 55, 3597-3610.

Li, H., Wickens, C. D., Sarter, N., & Sebok, A. (2014). Stages and levels of automation in support of space teleoperations. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 56(6), 1050-1061.

Metzger, U., & Parasuraman, R. (2005). Automation in future air traffic management: Effects of decision aid reliability on controller performance and mental workload. *Human Factors*, 47(1), 35-49.

Onnasch, L., Wickens, C. D., Li, H., & Manzey, D. (2014). Human performance consequences of stages and levels of automation: An integrated meta-analysis. *Human Factors*, 56(3), 476-488.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans*, 30(3), 286-297.

Perrow, C. (1984). *Normal accidents: Living with high risk systems*. New York: Basic Books.

Price, H. E. (1985). The allocation of functions in systems. *Human factors*, 27(1), 33-45.

Sarter, N. B., & Woods, D. D. (1997). Team play with a powerful and independent agent: Operational experiences and automation surprises on the Airbus A-320. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 39(4), 553-569.

Sarter, N. B., & Schroeder, B. (2001). Supporting decision making and action selection under time pressure and uncertainty: The case of in-flight icing. *Human factors*, 43(4), 573-583.

Sheridan, T. B. (2000). Function allocation: algorithm, alchemy or apostasy?. *International Journal of Human-Computer Studies*, 52(2), 203-216.

Sheridan, T. B., & Verplank, W. L. (1978). Human and computer control of undersea teleoperators. MASSACHUSETTS INST OF TECH CAMBRIDGE MAN-MACHINE SYSTEMS LAB.

Skraaning Jr., G., Hurlen, L., LeDarz, P., & Jamieson, G. A. (2016). Feasibility Study of an Interactive Large Screen Concept for Automated Plant Start-up (HWR-1179). OECD Halden Reactor Project.

Skraaning & Jamieson (in preparation). Twenty Years of HRP Research on Human-Automation Interaction: Insights on Automation Transparency and Levels of Automation (HPR-387). OECD Halden Reactor Project.

Society of Automotive Engineers (SAE; 2014). Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems (Standard J3016\_201401). Warrendale, PA: SAE International.