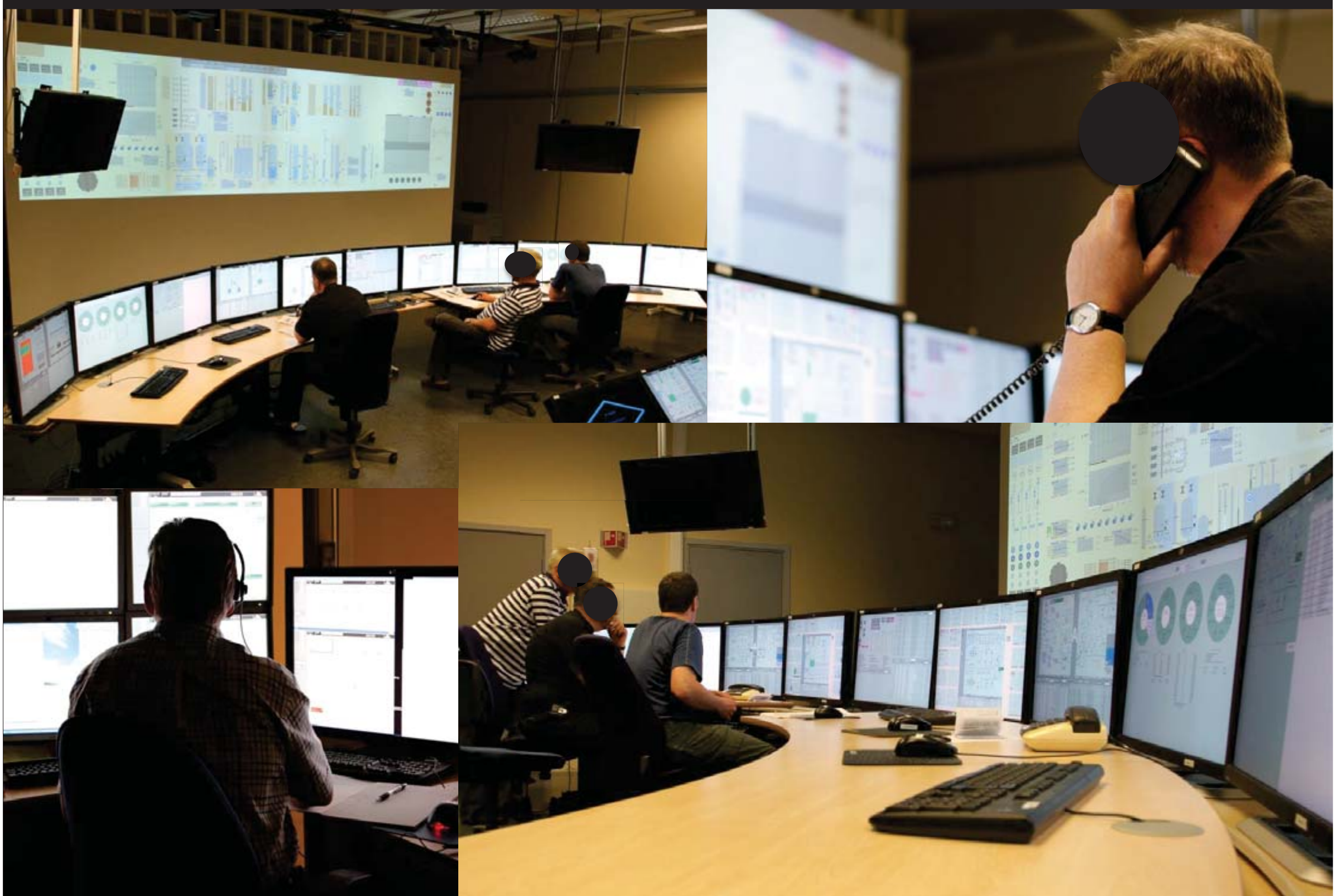
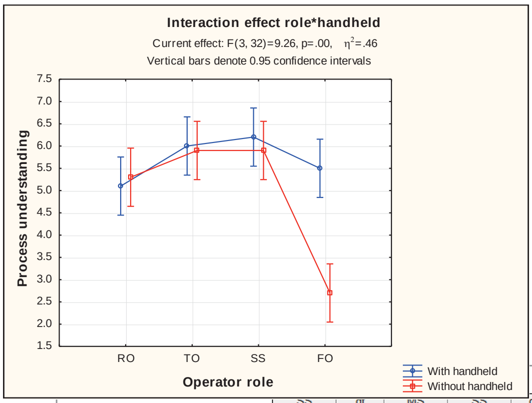
Apostasy in Level of Automation Modelling: Four Studies on the Road to Disillusionment

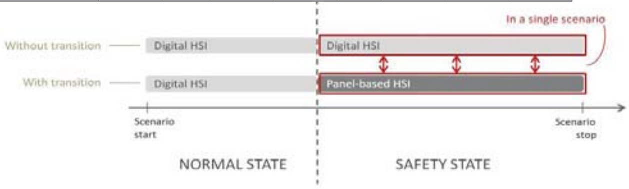
Greg A. Jamieson & Gyrd Skraaning Jr.

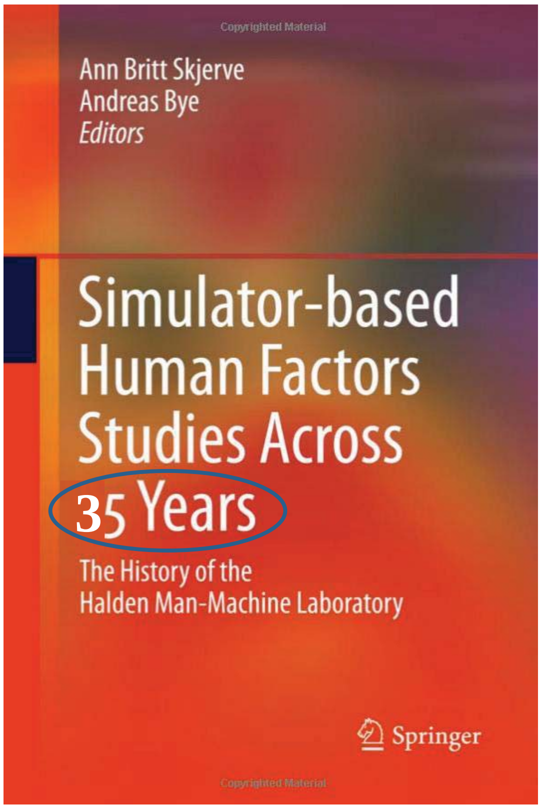
UNIVERSITY OF TORONTO
IFE HRP





Variable	SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error	F	p
OPAS task performance	1972.80	1.00	1972.80	3191.87	34.00	93.88	21.01	0.00
Process Expert rated task performance	2.92	1.00	2.92	11.47	34.00	0.34	8.66	0.01
Total communications per min	20.75	1.00	20.75	11.50	34.00	0.34	61.38	0.00
Self-rated period workload	1.21	1.00	1.21	15.54	34.00	0.46	2.65	0.11
Self-rated task performance	0.09	1.00	0.09	11.82	34.00	0.35	0.25	0.62
Self-rated situation awareness	0.22	1.00	0.22	16.53	34.00	0.49	0.46	0.50
Self-rated teamwork	0.06	1.00	0.06	15.02	34.00	0.44	0.13	0.73
Self-rated necessary of communication	0.15	1.00	0.15	15.28	34.00	0.45	0.34	0.56
Self-rated communication demand	0.08	1.00	0.08	20.30	34.00	0.60	0.13	0.72
Self-rated reliability of communication	0.14	1.00	0.14	7.24	34.00	0.21	0.65	0.42
Self-rated scenario workload	0.02	1.00	0.02	11.34	34.00	0.33	0.07	0.80





HAI studies

2018
Hierarchical HAI concept 2017 - FutureLab
LAD feasibility test -2015 - FutureLab

HAI prototyping 2011-2017 - FutureLab

2010
Out-of-the-loop performance 2009 - HAMMLAB
Coping with Automation 2009 - HAMMLAB

2005
Extended Teamwork 2004 - HAMMLAB

Procedure Automation 2002 - HAMMLAB
Procedure Automation 2001 - FITNESS
Human-Centered Automation 2001 - HAMMLAB
2000
Human-Centered Automation 2000 - HAMMLAB

Types of Automation 1998 - HAMMLAB
1997
Trust in Automation 1997 - HAMMLAB

HAI studies

2018
Hierarchical HAI concept 2017 - FutureLab
LAD feasibility test -2015 - FutureLab

HAI prototyping 2011-2017 - FutureLab

2010
Out-of-the-loop performance 2009 - HAMMLAB
Coping with Automation 2009 - HAMMLAB

2005
Extended Teamwork 2004 - HAMMLAB

Procedure Automation 2002 - HAMMLAB
Procedure Automation 2001 - FITNESS
Human-Centered Automation 2001 - HAMMLAB
2000
Human-Centered Automation 2000 - HAMMLAB

Types of Automation 1998 - HAMMLAB
1997
Trust in Automation 1997 - HAMMLAB

LOA

The Explorative LOA Study – HAMBO 2009

The LOA Classification Study – NORS 1998

1st Procedure Automation Experiment – FITNESS 2002

2nd Procedure Automation Experiment – HAMBO 2002

Acknowledgements

- Principal investigators in original studies
  - Ann Britt Skjerve
  - Maren Helene Rø Eitrheim
  - Gisle Andresen
- Process experts, developers, laboratory technicians, managers and others

Our contributions

- Statistical reanalysis
  - verify prior results
  - extract new findings
  - discover patterns across studies
- Reinterpretation of results
  - per study and across studies
- Connected results to recent literature

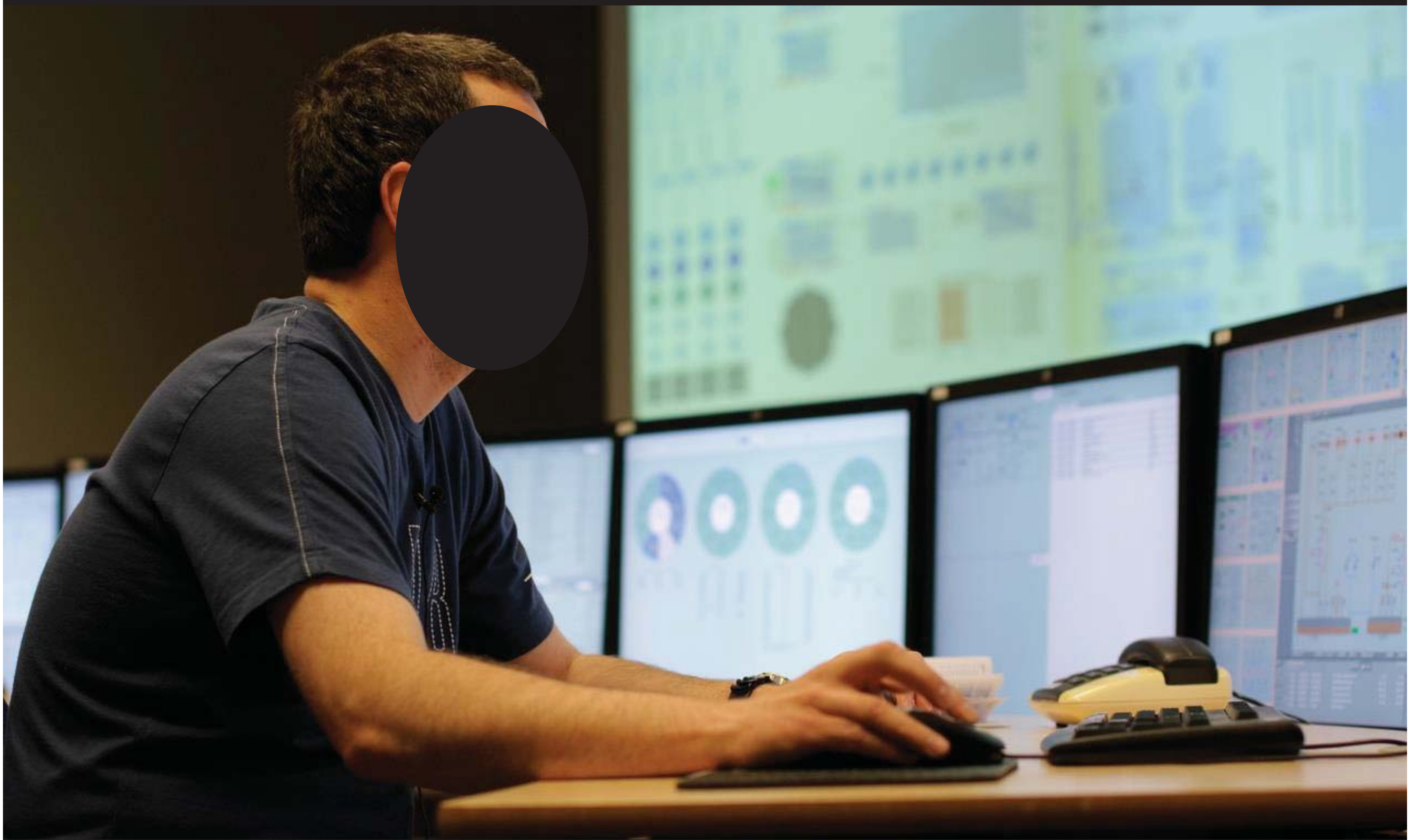
The Explorative LOA Study – HAMBO 2009

The LOA Classification Study – NORS 1998

1st Procedure Automation Experiment – FITNESS 2002

2nd Procedure Automation Experiment – HAMBO 2002

HAMMLAB 2009 – HAMBO Simulator



Complicated experiment

- Experimental manipulations
  - transparency of Automation (presented at HFC in 2014)
  - multi-unit staffing
  - overview displays
  - scenario complexity
- Explorative investigation
  - impact of high LOA on OOTLUF

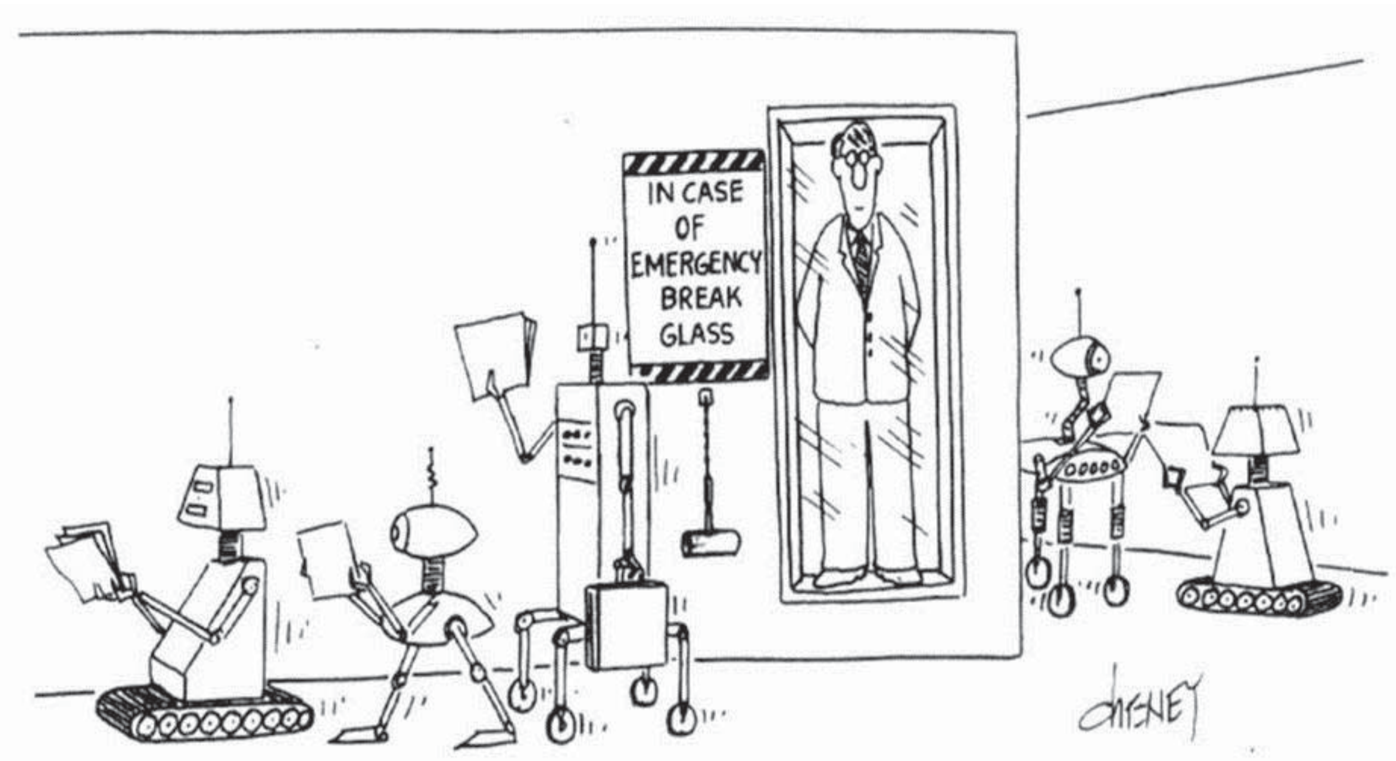
Run	Automation interface	Multi-unit staffing	Overview display
1	non-transparent	traditional	present
2	non-transparent	traditional	futuristic
3	transparent	traditional	futuristic
4	transparent	traditional	present
5	non-transparent	untraditional	present
6	non-transparent	untraditional	futuristic
7	transparent	untraditional	futuristic
8	transparent	untraditional	present

Easy scenario period

Difficult scenario period

Background

- When automatic systems are given more authority, intelligence and autonomy
  - humans are placed outside the control loop
  - but still expected to perform important functions, e.g. taking over manually in the event of system failures
- OOTLUF can be seen a consequence of increasing LOA



High LOA in the 2009 experiment

- Automation
  - executed process control tasks to bring the plant from cold start-up to 50% reactor power
  - performed basic monitoring of process deviations in normal system states
- Operators
  - supported automation when system failures occurred
  - were unfamiliar with automation that took active control of the whole plant

Multi-stage filtration

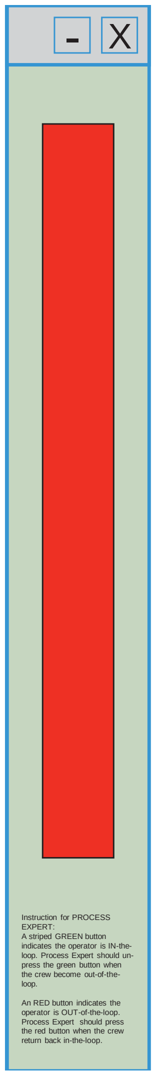
All human performance data

Possible OOTL events

Probable OOTL events

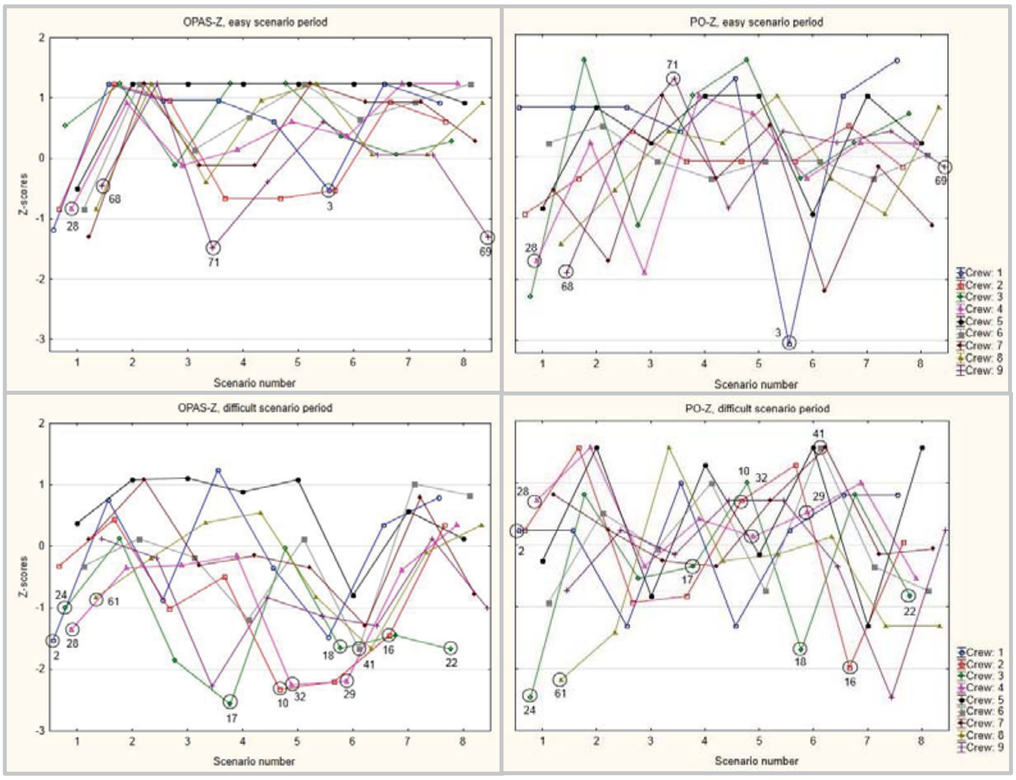
True OOTL events

OOTL indicators



- Loss of Situation Awareness
- Degraded Task Performance
- Out-of-the-loop buttons
- Video and simulator logs
- Debriefing interviews

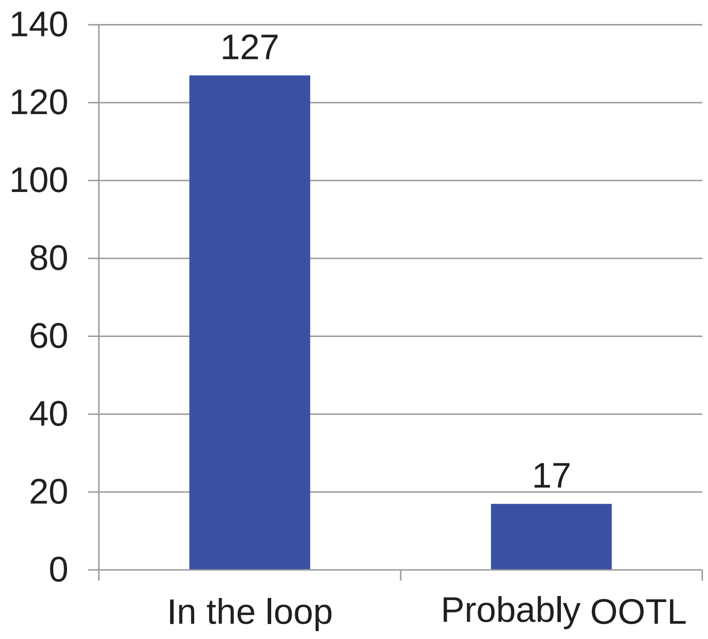
Searching for outliers



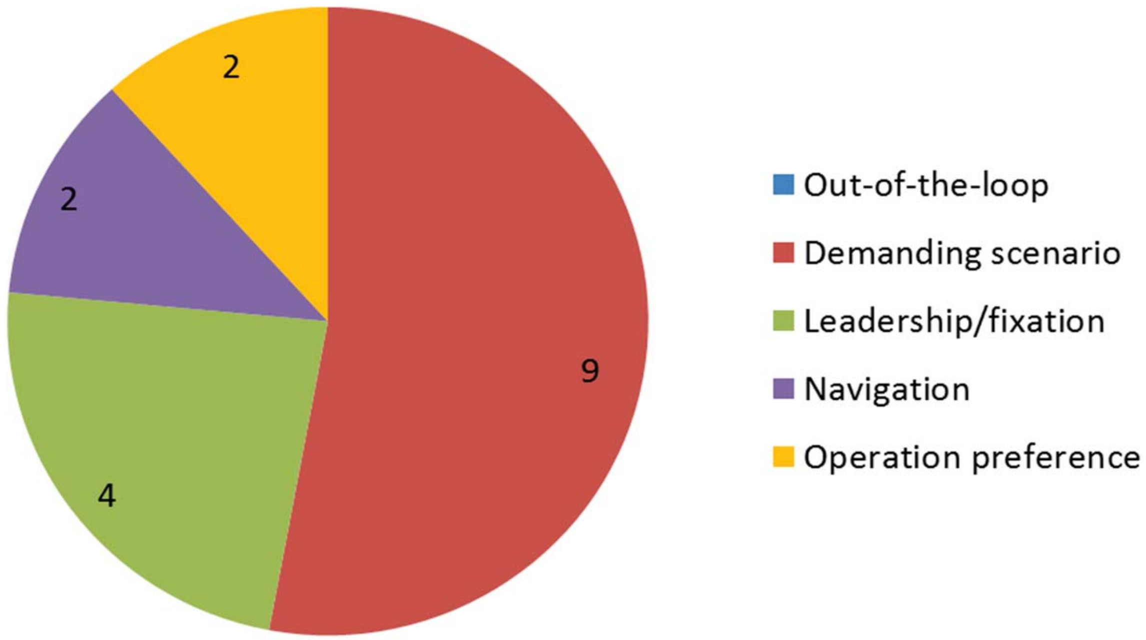
SME filtering of possible OOTL events

Scenario period	SCN	RUN	CREW	SUI	OPAS-Z	OPAS evaluation	PO-Z	PO evaluation	Experimental leader log	Process expert 1 log	Process expert 2 log	Filtering criteria
Easy	1	28	4	off	-0,84	Low score, but congruency between crews	-1,70	Remarkable low score	04.59 WM reminds MO that he has to take care of the turbine side as well as the reactor side			OPAS + PO
Easy	3	71	9	on	-1,48	Remarkable low score	1,29		MO and WM focused on 314, which has no relevance in this scenario. They're not detecting incoming alarms. Manual mode most of the period.	Spend 25 min on a system that has no relevance in the scenario.	A textbook example of what happens when both SS and the operator are focused on one problem, and loosing the overview of the process completely. It went totally wrong.	OPAS + SUI + logs
Easy	6	3	1	off	-0,52	Remarkable low score	-3,05	Remarkable low score	Navigation and technical questions - the crew got assistance from PE.	Did not manage to find appropriate picture for starting a redundant pump. The automation was put in manual for a long period (after break?).		OPAS + PO
Difficult	5	32	4	on	-2,25	Remarkable low score	0,13		The operators did not detect the reason for failures > reactor scram and turbine trip	Do not detect that condensate pump stopped - reactor scram, turbine trip.		OPAS + SUI + logs
Difficult	6	18	3	off	-1,66	Low score, but congruency between crews	-1,70	Remarkable low score	Some navigation/process questions to PE. Manual operation in the last part of period 1. Did not get the last error according to spec.	Detected increased flow, but did not solve the problem with air leakage to the condensor.		OPAS + PO
Difficult	6	41	6	on	-1,66	Low score, but congruency between crews	1,58		Are operators aware of automation actions? Operators are working as they would do at home, not aware of the automation progress	Reactor scram affects handling of error in the second period.	Tend to choose manual operation quite often. Working slowly and lack overview of the plant.	OPAS + SUI + logs
Difficult	7	16	2	off	-1,45	Remarkable low score	-1,99	Remarkable low score	AO in plant B the whole scenario, WM in plant B most of the time. MO navigation question.			OPAS + PO
Difficult	8	22	3	off	-1,67	Remarkable low score	-0,83		Manual operation during period 2. AO in plant B during the whole scenario	Did not detect bypass of preheaters.		OPAS + logs

17 probable OOTL events identified



SME analysis of 17 probable OOTL events



23

Reactions


The Explorative LOA Study – HAMBO 2009


The LOA Classification Study – NORS 1998


1st Procedure Automation Experiment – FITNESS 2002


2nd Procedure Automation Experiment – HAMBO 2002

HAMMLAB 1998 – NORS Simulator

Partly funded by IPSN (IRSN), France

Fundamental assumption behind the experiment

- LOA and taskwork are deeply entangled in real life
- Experimental separation of LOA and task effects creates artificial results
- LOA becomes an abstraction rather than a real phenomenon

Methodological approach

Population of scenarios

Scenario Scenario Scenario++
Scenario Scenario Scenario++
Scenario Scenario Scenario++
Scenario Scenario Scenario++
Scenario Scenario Scenario++

LOA in scenarios

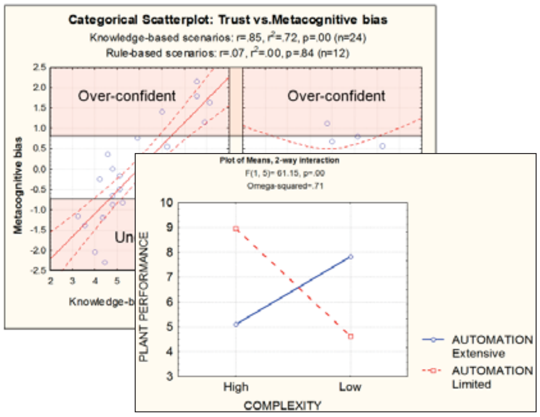
		High	Low
Scenario complexity	High	Scenario 1	Scenario 3
	Low	Scenario 4	Scenario 2

Adjustments to scenarios for perfect fit

A double-edged sword

- Prioritized ecological over internal validity
- LOA manipulations were deliberately confounded with scenario effects
- Challenging to interpret experimental effects
- Reanalysis produced ambiguous results

Scenario	Scenario type	Level of automation	Scenario Complexity
1	Diagnostic	High	High
2	Diagnostic	Low	Low
3	Diagnostic	Low	High
4	Diagnostic	High	Low
5	Procedural	High	Medium
6	Procedural	Low	Medium



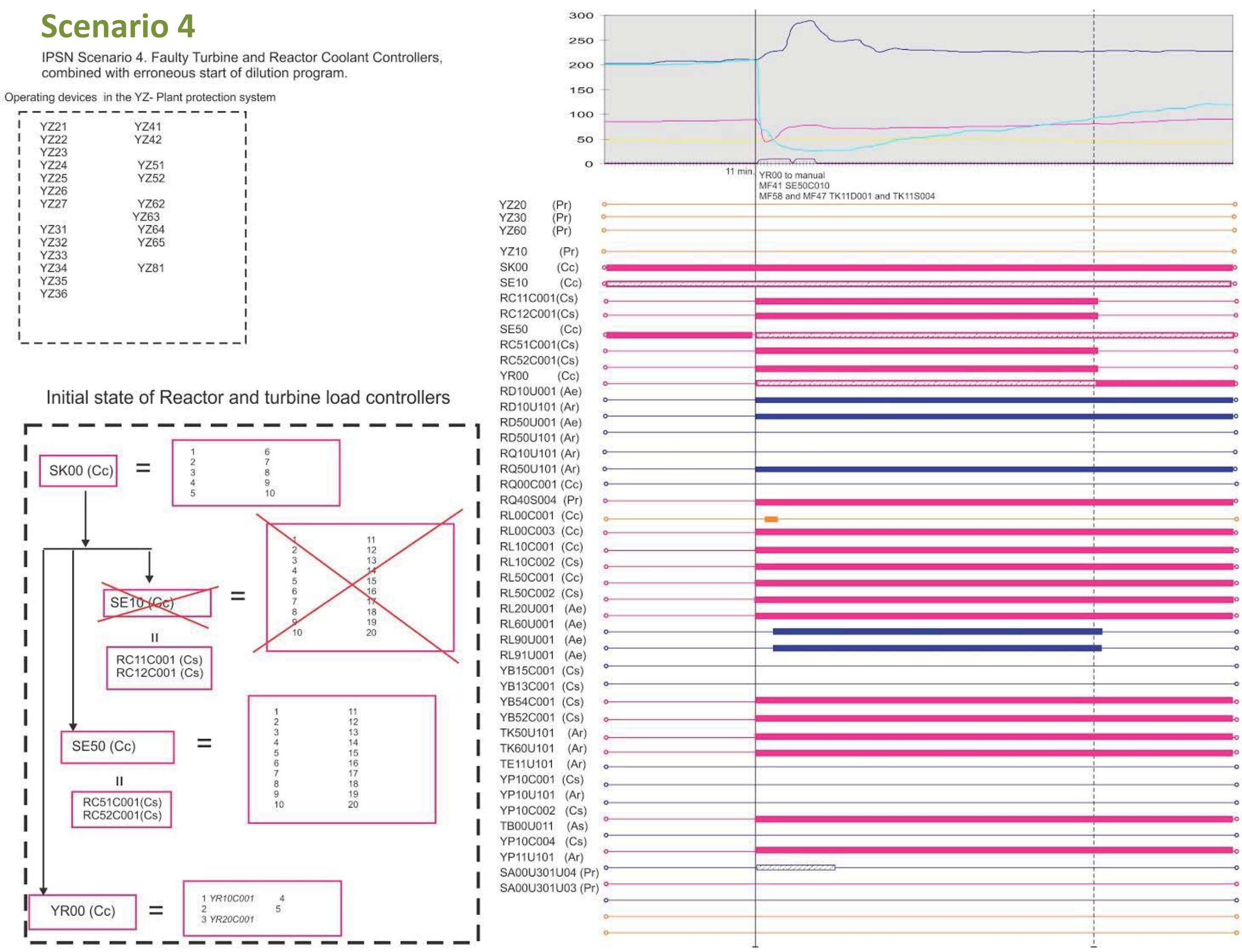
Substantiation of LOA classifications

- Six realistic test scenarios developed and classified by a team of SMEs
  - two process experts and a nuclear engineer
  - could they be wrong?

- We tried to verify their classifications or reclassify based on technically specific LOA criteria
- To improve the interpretability of the experiment

Scenario 4

IPSN Scenario 4. Faulty Turbine and Reactor Coolant Controllers, combined with erroneous start of dilution program.



31

Automatic devices in NORS

- **Interlocks** – prevent actions from being carried out in safety critical situations
- **Controllers** – regulate plant components (e.g. power level, temperature, pressure)
- **Limitations** – ensure that predefined operating values are not exceeded
- **Protections** – ensure that predefined safety critical values are not exceeded
- **Programs** – switch a group of components on or off in predefined sequences

Available information in 6 scenarios

	Number of involved	Number of active	Accumulated duration
Interlocks			
Controllers	✓	✓	✓
Limitations			
Protections	✓	✓	✓
Programs	✓	✓	✓
Manual replacements	✓		✓

90%

Home Layout Tables Charts SmartArt Formulas Data Review

F10 fx =(B10*12.2)-E10-(9.6*4)

	A	B	C	D	E	F
1	Scenario	Number of involved programs	Number of operating programs		Accumulated program operating duration	Accumulated program supervising duration
2	1	16	15	94	53.7	141.5
3	2	16	13	81	14.3	160.1
4	3	7	1	14	6.9	78.5
5	4	16	5	31	40.2	153.5
6	5	20	6	30	44.8	195.2
7	6	17	8	47	65.1	142.3
8						
9	Scenario	Number of involved controllers	Number of operating controllers		Accumulated controller operating duration	Accumulated controller supervising duration
10	1	22	20	91	147.6	82.4
11	2	22	17	77	92.2	120.2
12	3	15	11	73	86.3	63.3
13	4	21	19	90	161.6	66.8
14	5	22	15	68	158.9	80.9
15	6	22	19	86	169.8	87
16						
17						
18	Scenario	Number of involved protections	Number of operating protections	Accumulated protection operating durations	Accumulated protection supervising durations	
19	1	7	7	0.8	33.7	
20	2	7	5	1	48.5	
21	3	3	0	0	36.3	
22	4	7	1	0.2	84.5	
23	5	7	0	0	84.7	
24	6	7	1	0.3	75.5	
25						
26						
27	Scenario	Manual replacement of programs	Manual replacement of program duration	Manual replacement of controllers	Manual replacement of controller duration	
28	1	0	0	0	0	
29	2	0	0	0	0	
30	3	0	0	4	33.4	
31	4	1	1.5	3	27.8	
32	5	2	4	5	28.6	
33	6	0	0	2	11.6	
34						
35	Scenario	Auto active (sec)	Sum of devices	Sum operating devices	Duratin of manual effort	LOA
36	1	460	45	42	0	42
37	2	436	45	35	0	35
38	3	271	25	12	33.4	8
39	4	507	44	25	29.3	21
40	5	565	49	21	32.6	14
41	6	540	46	28	11.6	26
42						

Examples of tested LOA indexes

- Σa_i(controller, protection and automatic program activities) - Σb_i(manual replacements)
- Accumulated duration of [controller, protection and automatic program activities] – accumulated duration of manual efforts
- Combinations of duration and frequency indicators
- Automation activity without taking manual replacements into account
- Separate indexes for controllers, protections and automatic programs

Outcome

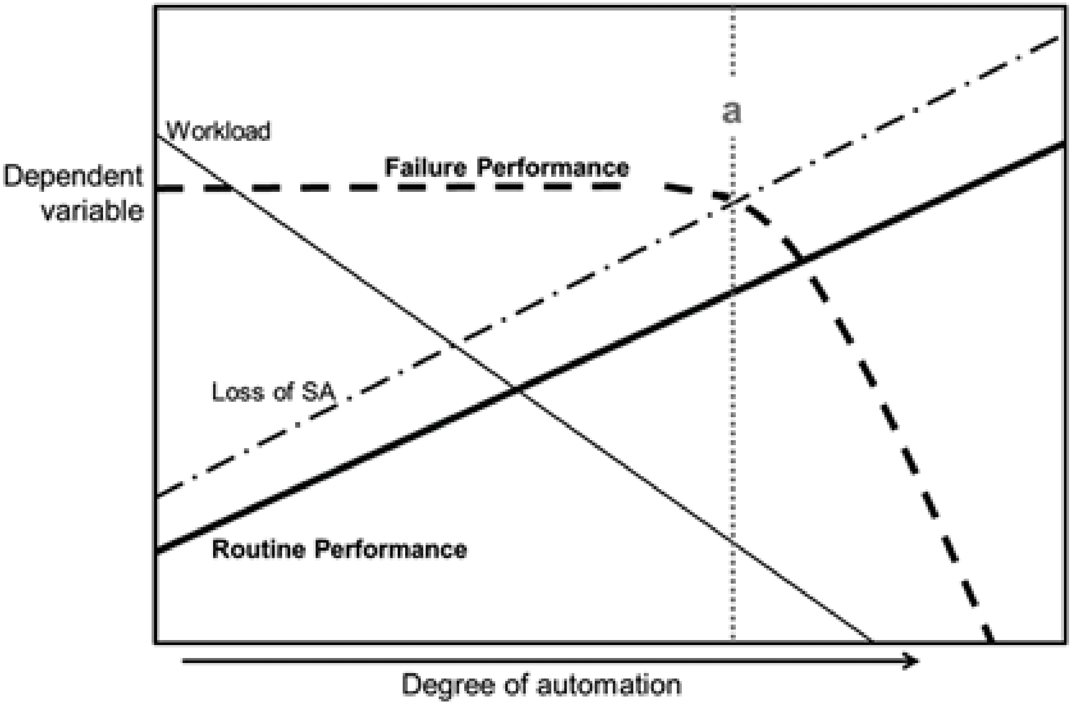
- Many disparate calculations
- Inability to match calculations to original SME assessment
- Almost any LOA classification of scenario was possible
- Pessimistic interpretation
  - LOA and taskwork seem inseparable in real life
  - LOA classification of complex taskwork appears problematic
  - is LOA an abstraction without root in reality?

37

From country cruising to Rally Dakar

>

Level (or degree) of automation



39

Adapted from Wickens et al. (2010) by Onnasch et al. (2014)

Human Performance Consequences of Stages and Levels of Automation: An Integrated Meta-Analysis

Linda Onnasch, Technische Universität Berlin, Berlin, Germany,
Christopher D. Wickens, Alion Science and Technology, McLean, Virginia, USA, **Huiyang Li**, University of Michigan, Ann Arbor, Michigan, USA, and
Dietrich Manzey, Technische Universität Berlin, Berlin, Germany

HUMAN FACTORS
Vol. 56, No. 3, May 2014, pp. 476–488
DOI: 10.1177/0018720813501549

TABLE 1: Kendall's Tau for the Single Studies on the Six Metavariables With Resulting Overall Kendall's Tau and Statistics of One-Tailed *t* Tests

Study	Routine Primary Task Performance (TP)	Return-to-Manual Primary TP	Routine Secondary TP	Return-to-Manual Secondary TP	Subjective Workload	Situation Awareness
Calhoun et al. (2009)	−.816		0			0
Crocoll & Coury (1990)	.707					
Cummings & Mitchell (2007)	0					0
Endsley & Kaber (1999)	.637	.025			.804	.597
Endsley & Kiris (1995)		−.837			0	−.837
Kaber & Endsley (2004)	.6	0			−.598	.258
Kaber et al. (2000)	.316	−.408			−.775	−.632
Li et al. (in preparation)	1				−1	−1
Lorenz et al. (2002a)	.333	−.333	0	0	0	
Lorenz et al. (2002b)	.816	.333				
Manzey et al. (2012)	.913	−.816	.913		−.913	−.707
Metzger & Parasuraman (2005)	0	0	0	0	0	0
Reichenbach et al. (2011)	1	−1	0	0	0	0
Röttger et al. (2009)	.816		0		−1	
Rovira et al. (2007)	.837		.707		−.333	
Sarter & Schroeder (2001)	1					
Sethumadhavan (2009)			.707			−.913
Wright & Kaber (2005)	0				.913	
Overall τ	.509	−.337	.291	0	−.242	−.294
t-crit	1.341	−1.397	1.415		−1.363	−1.372
t	4.027	−2.176	2.024		−1.284	−1.809
p	.0005*	.031*	.042*		.056	.049*

- clear automation benefit for routine performance with increasing DOA
- similar but weaker pattern for workload
- negative impact of higher DOA on failure system performance and SA

HUMAN FACTORS
Vol. 56, No. 3, May 2014, pp. 476–488
DOI: 10.1177/0018720813501549


The Explorative LOA Study – HAMBO 2009

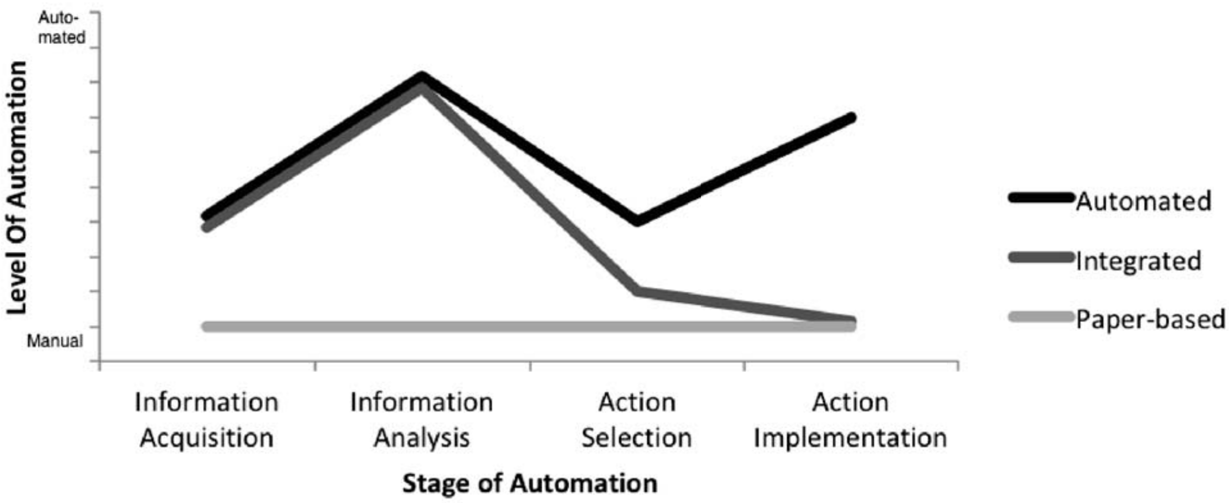

The LOA Classification Study – NORS 1998


1st Procedure Automation Experiment – FITNESS 2002

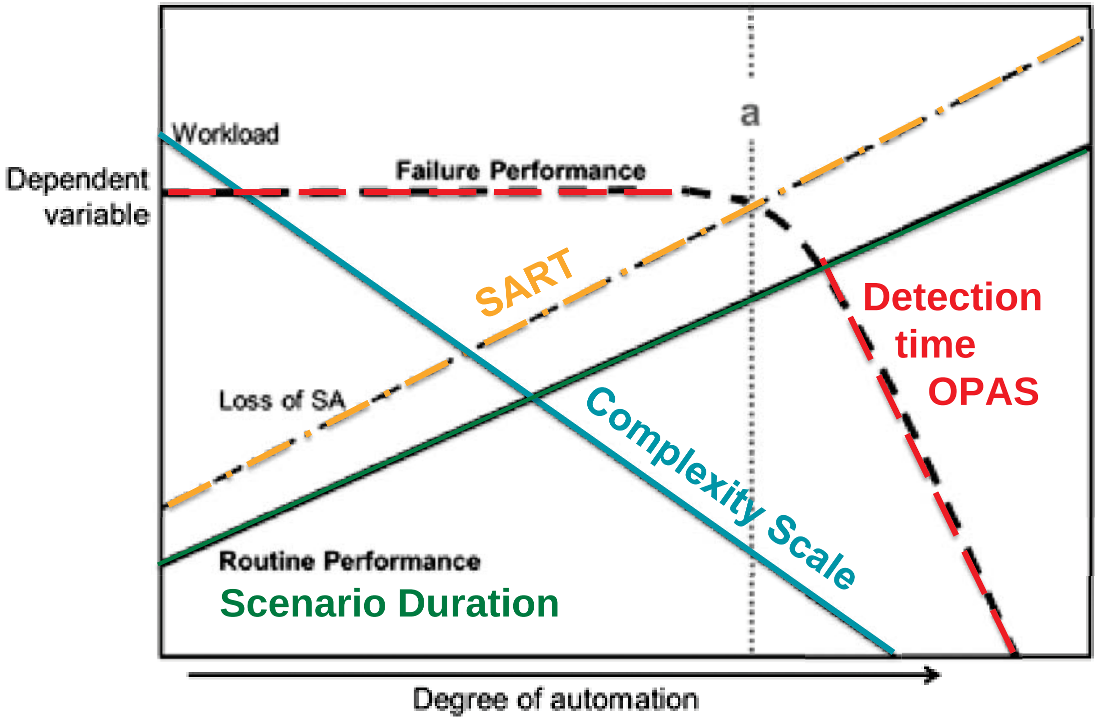

2nd Procedure Automation Experiment – HAMBO 2002

EDF SEPTEN 2002 – FITNESS Simulator

Degree of procedure automation

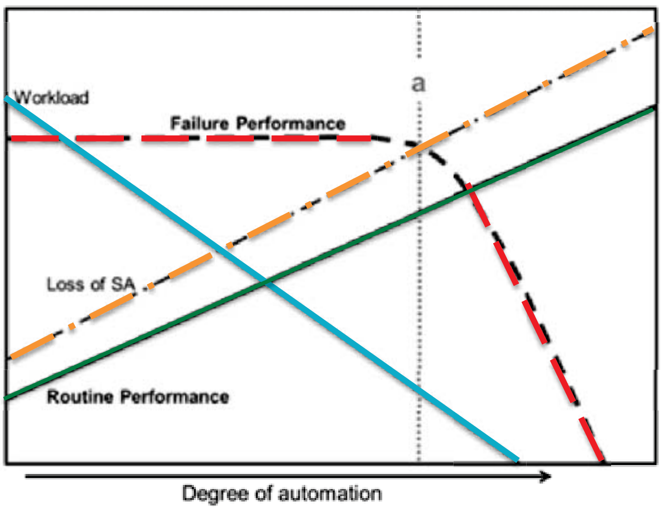


- Profiling based on
  - information-processing perspective (Parasuraman et al., 2000)
  - COPS capability analysis (IEEE Std. 1786-2011)



Statistical findings

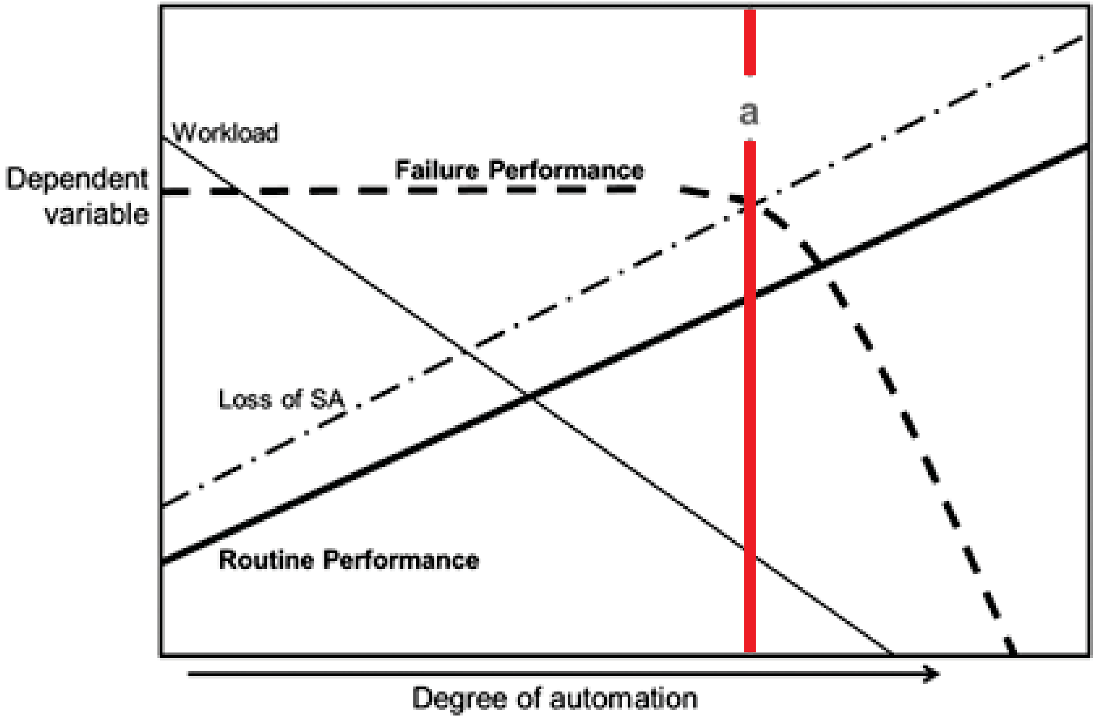
- Routine performance more efficient with automated procedures
  $F(2,10) = 6.80$, $p < 0.014$
- LOA had no effect on failure performance, situation awareness, or workload



Possible explanations

- Participants were mostly engineers rather than licensed operators
- Participants worked individually; not in crews
- Participants not required to use the COPS
- “to be able to measure the effect of procedure automation, more sensitive and valid measures are required” (HWR-707, p. 19)

Explained by the ‘a’ threshold?



 The Explorative LOA Study – HAMBO 2009

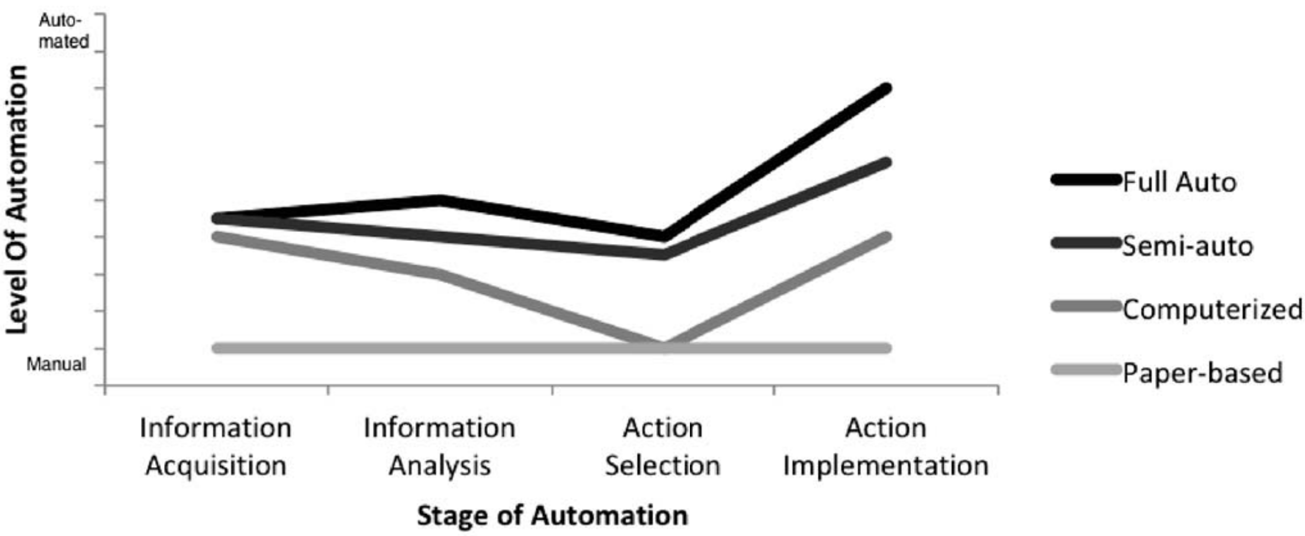
 The LOA Classification Study – NORS 1998

 1st Procedure Automation Experiment – FITNESS 2002

 2nd Procedure Automation Experiment – HAMBO 2002

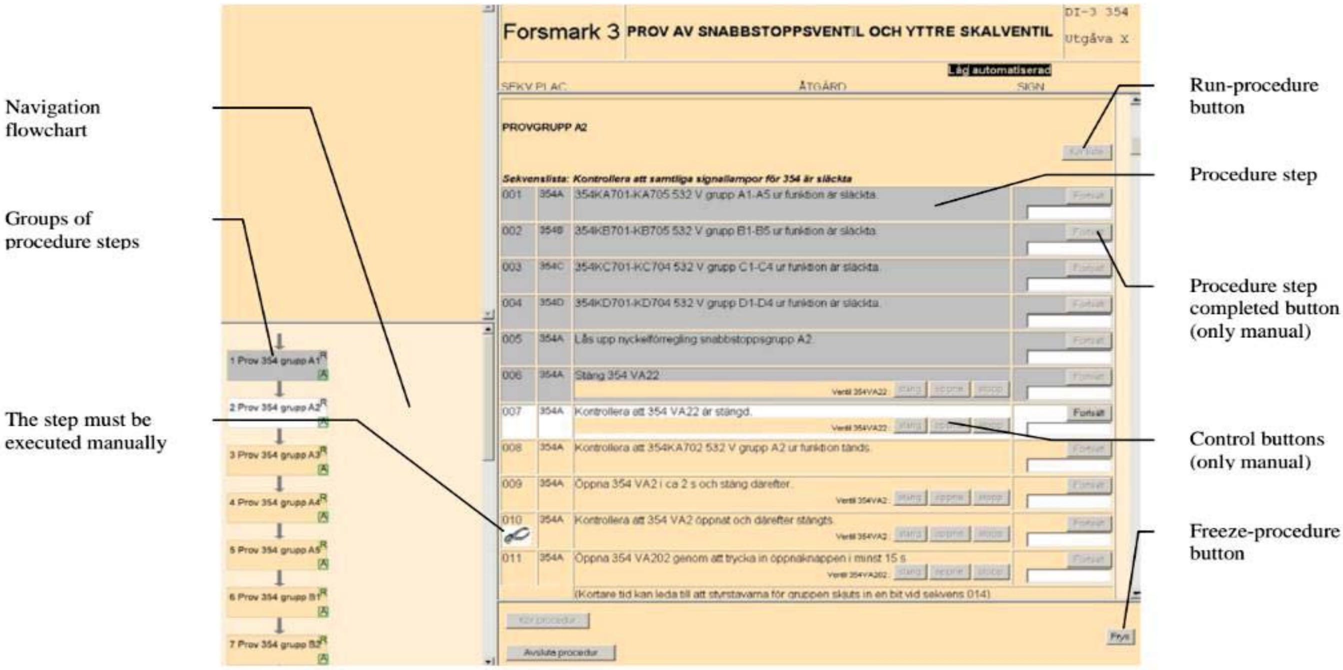
HAMMLAB 2002 – HAMBO simulator

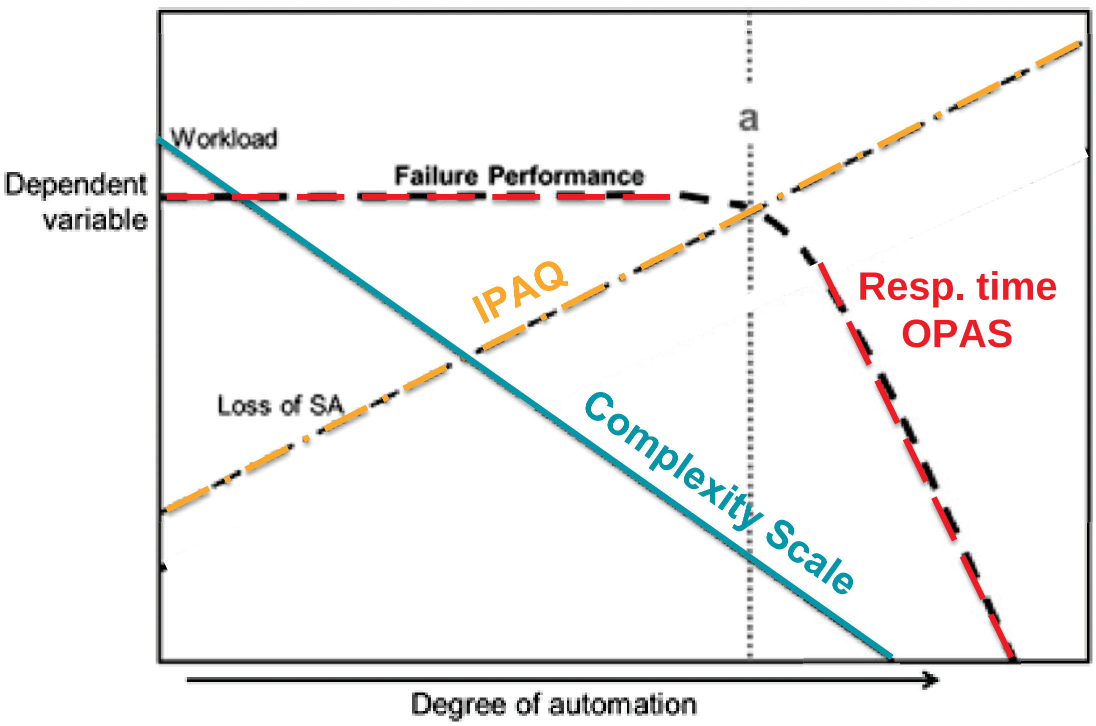
Degree of procedure automation

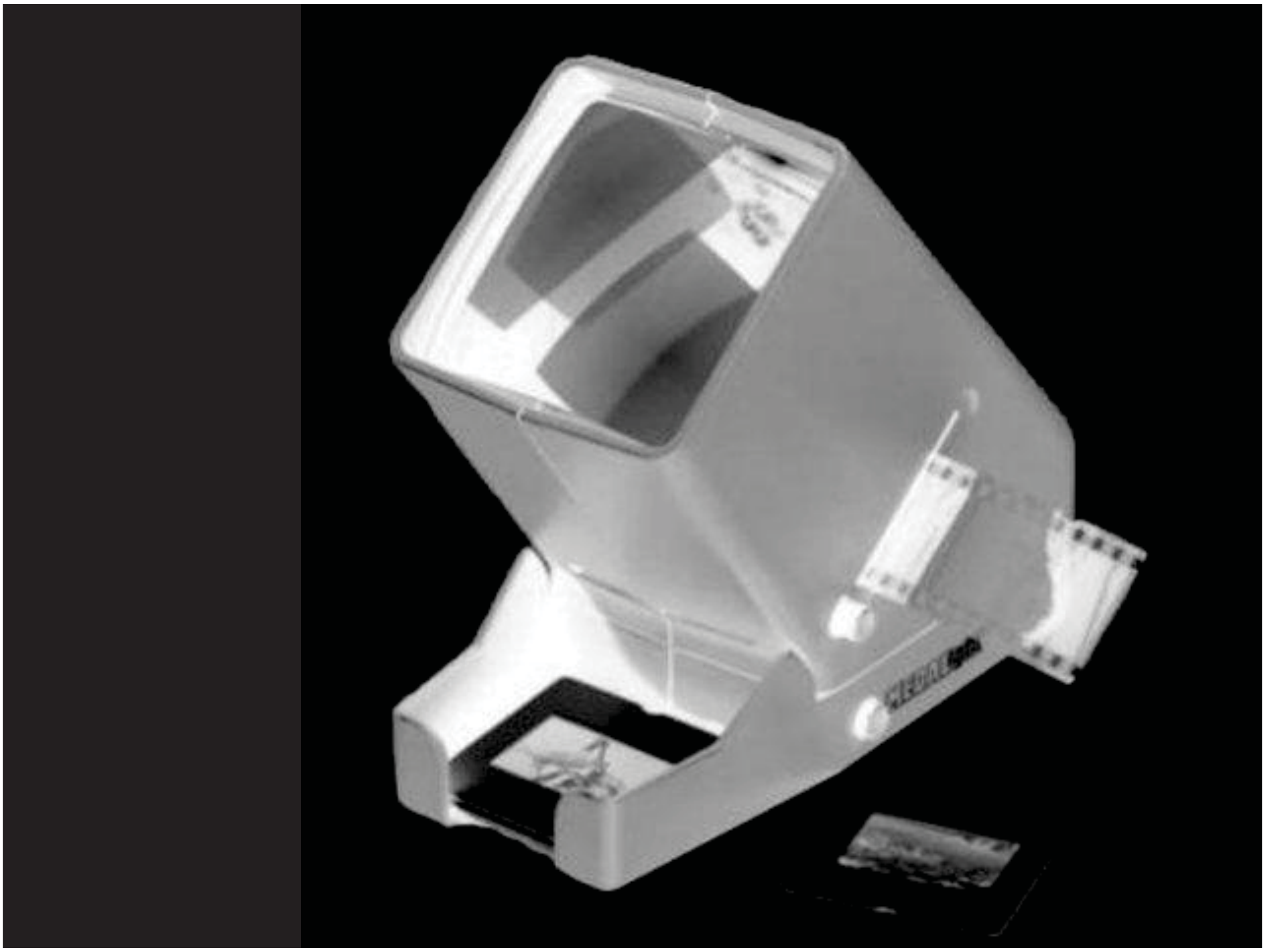


- Profiling based on
  - information-processing perspective (Parasuraman et al., 2000)
  - COPS capability analysis (IEEE Std. 1786-2011)

COPMA procedure system





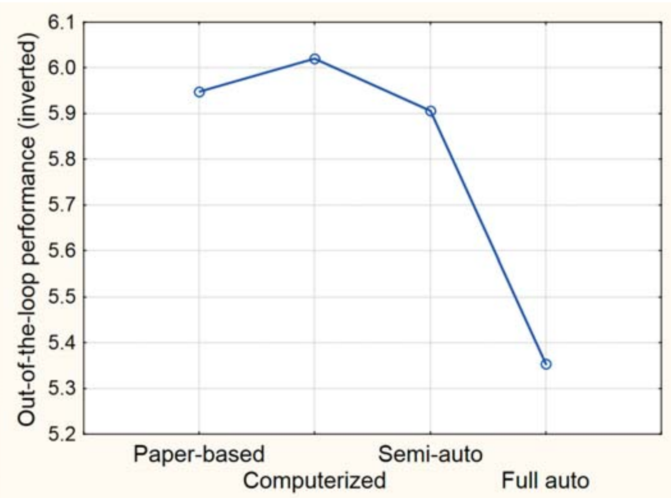


The 2nd Procedure
Automation Experiment

Support for lumberjack model

Out-of-the-loop performance

Direction of effect

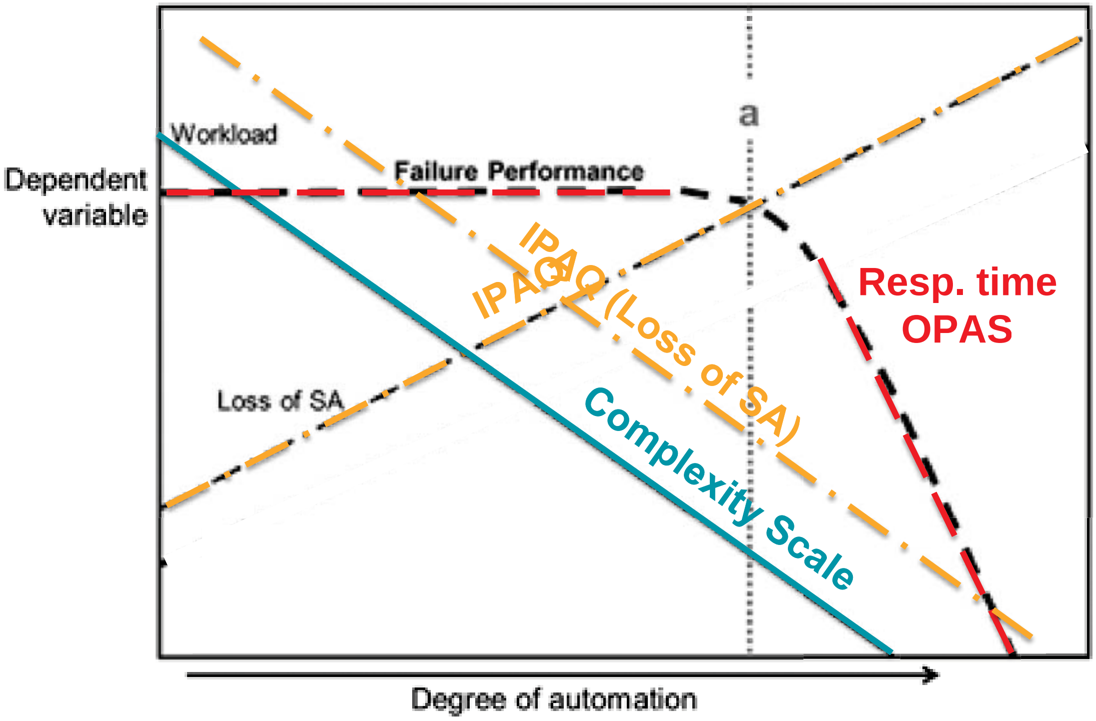


$F(3,63)=24.69$, $p=0.001$

$\eta^2=0.41$ (effect size)

Interpreted support

Higher scores mean less experienced
difficulty with the procedure execution



Dependent variable
Workload
Failure Performance
a
IPAQ
IPAQ (Loss of SA)
Resp. time OPAS
Loss of SA
Complexity Scale
Degree of automation

Possible explanations

- Critique
  - inflated error term due to scenario variation
  - unfamiliar SA measure (IPAQ)
- Response
  - high degree of similarity in methods to other HAMMLAB experiments that demonstrated anticipated effects
  - IPAQ sensitive to scenario manipulations; works well in HAMMLAB experiment on teamwork and task management (HWR-704)

Possible Explanations

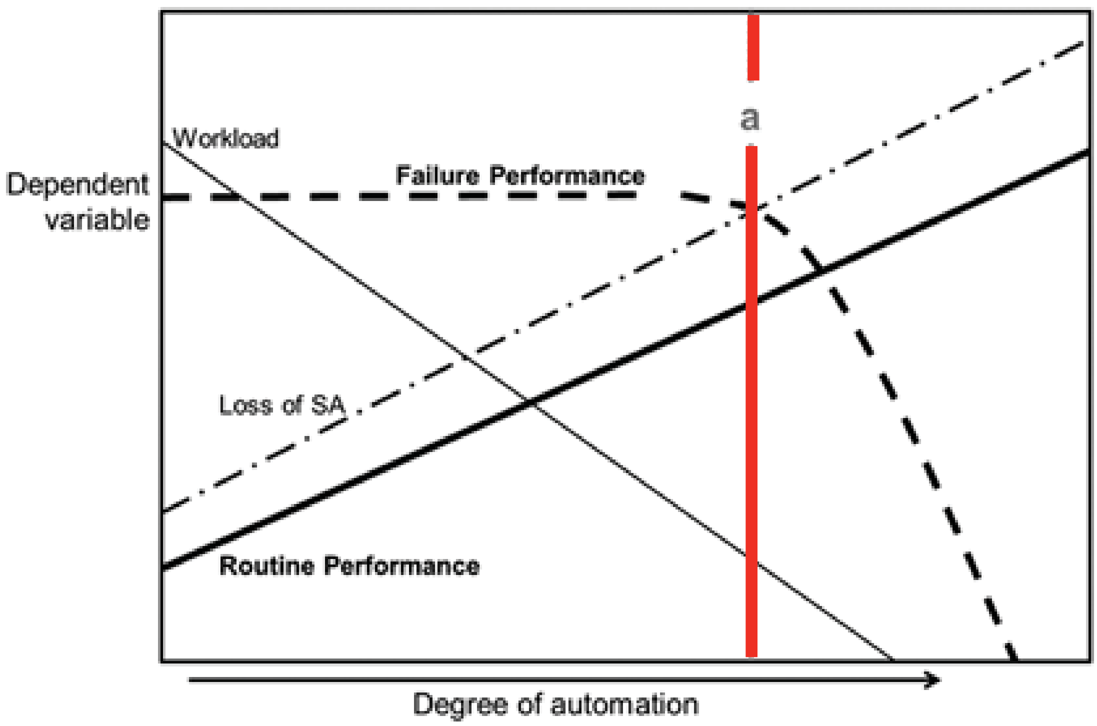


TABLE 1: Kendall's Tau for the Single Studies on the Six Metavariables With Resulting Overall Kendall's Tau and Statistics of One-Tailed *t* Tests

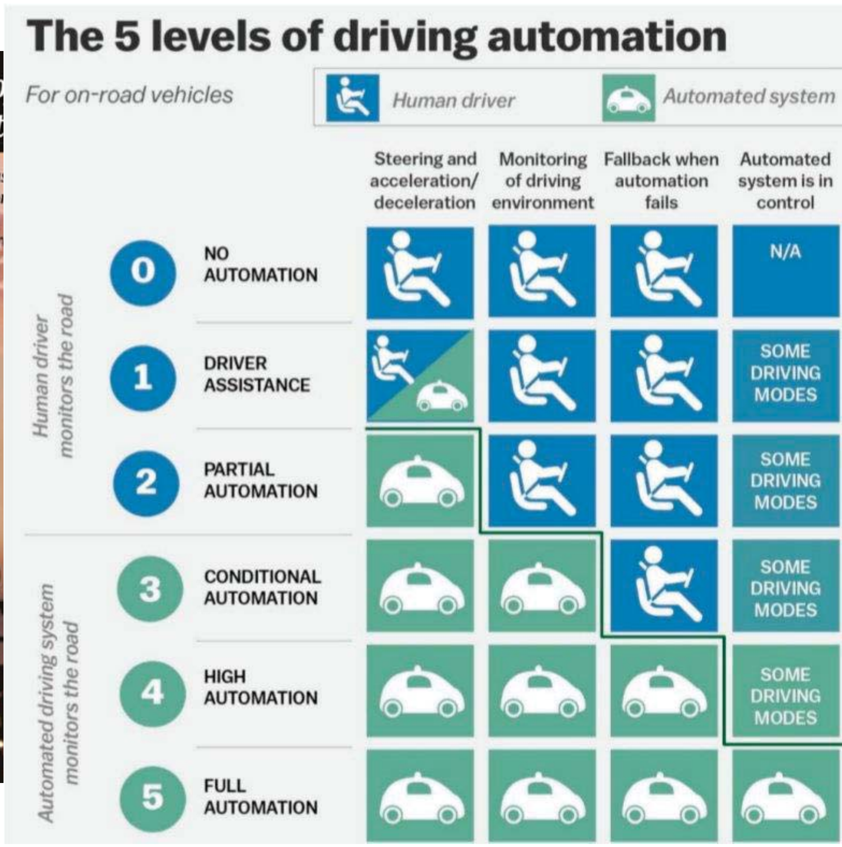
Study	Routine Primary Task Performance (TP)	Return-to-Manual Primary TP	Routine Secondary TP	Return-to-Manual Secondary TP	Subjective Workload	Situation Awareness
Calhoun et al. (2009)	–.816		0			0
Crocoll & Coury (1990)	.707					
Cummings & Mitchell (2007)	0					0
Endsley & Kaber (1999)	.637	.025			.804	.597
Endsley & Kiris (1995)		–.837			0	–.837
Kaber & Endsley (2004)	.6	0			.598	.258
Kaber et al. (2000)	.316	–.408			–.775	–.632
Li et al. (in preparation)	1				–1	–1
Lorenz et al. (2002a)	.333	–.333	0	0	0	
Lorenz et al. (2002b)	.816	.333				
Manzey et al. (2012)	.913	–.816	.913		–.913	–.707
Metzger & Parasuraman (2005)	0	0	0	0	0	0
Reichenbach et al. (2011)	1	–1	0	0	0	0
Röttger et al. (2009)	.816		0		–1	
Rovira et al. (2007)	.837		.707		–.333	
Sarter & Schroeder (2001)	1					
Sethumadhavan (2009)			.707			–.913
Wright & Kaber (2005)	0				.913	
Overall τ	.509	–.337	.291	0	–.242	–.294
t-crit	1.341	–1.397	1.415		–1.363	–1.372
t	4.027	–2.176	2.024		–1.284	–1.809
p	.0005*	.031*	.042*		.056	.049*

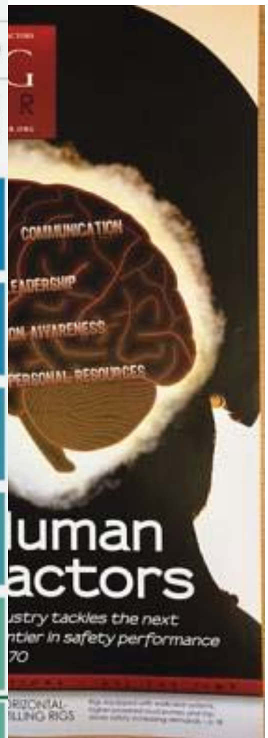
- clear automation benefit for routine performance with increasing DOA, similar but weaker pattern for workload
- **Complex work settings**
  - professional operators
  - high fidelity simulators
  - realistic scenarios
- negative impact of higher DOA on failure system performance and SA.

HUMAN FACTORS
Vol. 56, No. 3, May 2014, pp. 476–488
DOI: 10.1177/0018720813501549

Study	Experiment Characteristics	Routine Primary Task Performance	Return-to-Manual Primary Task Performance	Routine Secondary Task Performance	Return-to-Manual Secondary Task perform	Subjective Workload	SA
Calhoun et al. (2009)	Military personnel (non-SME); Commercial simulator; complex scenarios	-.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 & Schroeder (2001)	Commercial aircraft pilots; Full-scope simulator; Complex scenarios	1					
1st Procedure Automation Experiment	Licensed NPP operators; Full- scope simulator; Complex scenarios	1	0			0	0
2nd Procedure Automation Experiment	Licensed NPP operators; Full- scope simulator; Complex scenarios	0				0	1

Operational Experience with LOA





A consistent pattern of results

- Explorative studies
  - anticipated OOTL performance degradations under high LOA were absent
  - technically specific LOA classification of realistic test scenarios produced arbitrary results
- Findings from two controlled LOA experiments were inconsistent with the lumberjack model
  - compared to similar studies from complex work settings in other domains
  - compared to operational experiences with LOA applications

Practical implications

- Technical basis of LOA predictions in complex work settings seems weaker than anticipated
- Be critical when applying established LOA frameworks
- LOA modeling could mislead designers and result in unsafe human-machine systems
- Ineffective LOA regimes may undermine the legitimacy of human factors design guidance

jamieson@mie.utoronto.ca
gyrds@ife.no