2022 IEEE 3rd International Conference on Human-Machine Systems (ICHMS) | 978-1-6654-5238-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICHMS56717.2022.9980609

Human operators' blind compliance, reliance, and dependence behaviors when working with imperfect automation: A meta-analysis

Patrik T. Schuler
University of Michigan
Department of Industrial and Operations Engineering
Ann Arbor, USA
Patriks@umich.edu

X. Jessie Yang
University of Michigan
Department of Industrial and Operations Engineering
Ann Arbor, USA
Xijyang@umich.edu

***Abstract*—We conducted a meta-analysis to determine how people blindly comply with, rely on, and depend on diagnostic automation. We searched three databases using combinations of human behavior keywords with automation keywords. The period ranges from January 1996 to June 2021. In total, 8 records and a total of 68 data points were identified. As data points were nested within research records, we built multi-level models (MLM) to quantify relationships between blind compliance and positive predictive value (PPV), blind reliance and negative predictive value (NPV), and blind dependence and overall success likelihood (OSL).**

Results show that as the automation's PPV, NPV, and OSL increase, human operators are more likely to blindly follow the automation's recommendation. Operators appear to adjust their reliance behaviors more than their compliance and dependence. We recommend that researchers report specific automation trial information (i.e., hits, false alarms, misses, and correct rejections) and human behaviors (compliance and reliance) rather than automation OSL and dependence. Future work could examine how operator behaviors change when operators are not blind to raw data. Researchers, designers, and engineers could leverage understanding of operator behaviors to inform training procedures and to benefit individual operators during repeated automation use.

***Index Terms*—human-automation interaction, automation use, reliance, compliance, dependence**

I. Introduction

Since Parasuraman and Riley's seminal paper on automation use [1], there is an increasing amount of research examining the influence of automation on human behaviors and task performance [2]–[4]. Automation provides many benefits, including improved operator safety, reduced costs, and reduced operator workload. However, when automation is not 100% reliable, rather than enhancing team performance, automation could instead deteriorate it due to inappropriate use of automation [2]. Specific incidents and accidents due to under- or over-utilization of automation have been reported. However, we do not know, *on average*, how people blindly follow automation, especially as automation performance varies.

This research was funded in part by the National Science Foundation under Grant No. 2045009.

In the present study, we conduct a meta-analysis to determine how people blindly comply with, rely on, and depend on diagnostic automation. The scope of the meta-analysis is limited to diagnostic automation. According to the four-stage taxonomy of automation [5], diagnostic automation refers to stage 2 automation, wherein the automation infers the state of the world, either "signal present" (i.e., an alert is provided) or "signal absent".

Compliance, Reliance, and Dependence Behaviors

If automation were perfect, the operator's decision would be simple: follow the automation's recommendation. When automation is imperfect, the operator's decision becomes more complicated, as they must decide whether to follow the automation. When working with diagnostic automation, there exist two types of behaviors, namely compliance and reliance [6]. Compliance refers to the human operator's tendency to perform an action when the automation diagnoses a signal in the world, whether true or false, and reliance is the human operator's tendency to refrain from performing an action when the automation is silent, indicating "all is well" [6], [7].

In this study, we are interested in blind compliance and reliance behaviors – the probability that the operator follows the automation's recommendation *without* cross-checking the raw information. Blind compliance is calculated as *Pr(report AND not cross-checking |alert)*. Blind reliance is calculated as *Pr(not report AND not cross checking|no alert)*. Dependence refers to the human's behavior during both alerts and non-alerts; it is a pooled measure of reliance and compliance behaviors, calculated as *Pr(human blindly follows automation)*.

Predictive Values

Signal detection theory (SDT) can be used to measure diagnostic automation's performance [3], [8], [9]. This study focused on three SDT measures of automation performance: the positive predictive value (PPV), the negative predictive value (NPV), and the overall success likelihood (OSL) - OSL is commonly referred to as reliability. PPV is defined as the probability of a true signal given an automation alert, Pr(signal|alert), and NPV is defined as the probability of not having a signal given the automation is silent, Pr(no signal|no alert). OSL refers to the automation's performance during both alerts and non-alerts; it was defined as the percentage of the time an automation aid is correct, Pr(correct automation).

II. Methods

We conducted a literature search to identify relevant studies, following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) process [10].

A. Relevant Literature Aggregation

A database search was conducted for records containing domain-specific term combinations. The databases included EBSCOhost, ProQuest, and IEEE Xplore. The combinations consisted of a key term from Set1 (focusing on automation) along with a key term from Set2 (focusing on human behavior). One example combination was "automation AND dependence".

$$Set1 = \{\text{"automation", "automated system", "autonomy"}\};$$
$$Set2 = \{\text{"dependence", "trust", "compliance", "reliance"}\}.$$

Term combinations were searched for within metadata, including abstract, keywords, and title. Search results were restricted to conference proceedings papers, journal publications, and thesis papers. If data were reported in a conference proceeding paper and thesis/publication, the thesis/publication was selected because more specific data were available. The record publishing date was restricted to a period ranging from January 1996 to June 2021, which was selected by bench-marking prior literature reviews [11]–[13]. Additionally, a language restriction was implemented to ensure that results were written in English.

The initial searches yielded 15,228 records. As multiple term combinations were used in various search engines, there were many duplicate records. Once duplicates were removed, 10,328 unique records remained.

Screening Process. Each of the 10,328 records was then screened according to the following criteria:

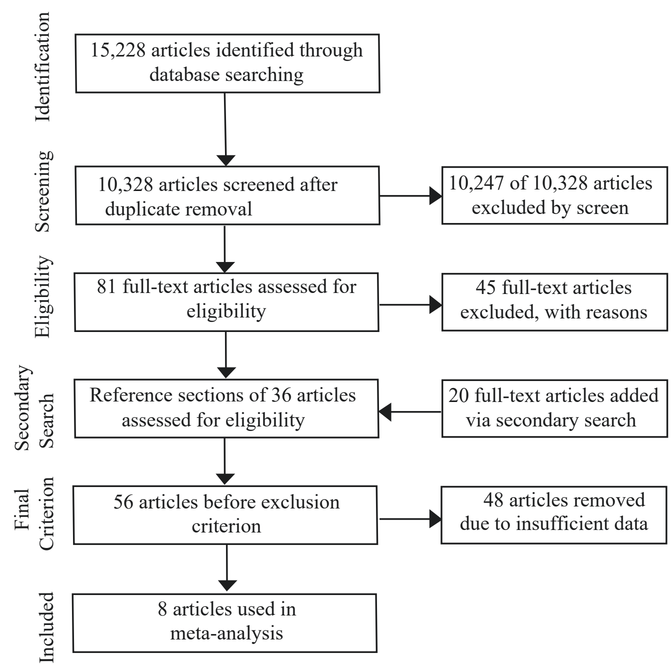
1) The record yielded original data from a human-subject experiment;
2) The human-subject experiment employed diagnostic automation, whose performance could be quantified using SDT;
3) At least one experimental condition used imperfect automation;
4) A dual-task paradigm was used in the experiment.

The screening process excluded 10,247 records. Most of these were excluded because the authors did not conduct a human-subject experiment.

Record Eligibility. The remaining 81 full-text records were carefully reviewed for eligibility. The eligibility criteria were:

1) Each record contained sufficient SDT data (i.e., # of hits, misses, false alarms, and correct rejections) to quantify automation performance (i.e., positive predictive value, negative predictive value, and automation overall success likelihood);
2) Empirical data reported in a manner that allows for calculations of compliance, reliance, or dependence (i.e., regardless of whether the compliance, reliance, or dependence is blind or not).

Secondary Search and Final Criterion. There were 36 records remaining after the first two steps of the exclusion process. Next, domain-specific records were sought out in the reference sections of the 36 records. This secondary search was conducted via Google Scholar, and a screening and eligibility process was conducted on every reference article. The secondary search yielded 20 additional records and increased the total count to 56. A final exclusion criterion ensured participants did not access the raw data, capturing blind reliance, compliance, and dependence, thus resulting in a total of 8 records in the analyses - these records are noted with an asterisk in the References section. The record qualification process is visualized below in Figure 1.



Flow Chart (ii).pdf

Fig. 1. The PRISMA flowchart illustrates the literature selection process.

B. Variable Calculation and Analyses

From the 8 records, the following metric combinations were extracted for analysis: (I) operator blind compliance and automation PPV, (II) operator blind reliance and automation NPV, (III) operator blind dependence and automation OSL. If not explicitly reported, dependence was calculated with reported compliance and reliance data.

Because data points were nested by record (one record could have multiple experiments that were not independent), we built multi-level models (MLM) to analyze the three relationships using the 'nlme' package in R (version 4.0.3). The automation's performance metrics (PPV, NPV, and OSL) were considered the predictors of human behaviors in each MLM. Following the standard procedure in building MLM [15], we gradually built more complex models, from the fixed-effects model to the random intercept model and finally to the random slope random intercept model. Using the Analysis of Variance (ANOVA) test, we compared the Akaike information criterion (AIC) scores between a simpler and a more complex model to determine whether the latter was needed. Whenever we found a non-significant comparison or a more complex model failed to converge, we reverted to a simpler model. The level of significance for this study was set to $\alpha = 0.05$.

III. Results

68 data points from 8 records reported operators' blind compliance. The reported blind compliance ranged from 0% to 98%, with a mean of 45.14% (SD = 25.38%). 66 data points from 7 records reported operators' blind reliance. The reported blind reliance ranged from 24% to 99%, with a mean of 64.14% (SD = 21.51%). The blind compliance rates were much lower than the blind reliance rates. 66 data points from 7 records contained operators' blind dependence, which ranged from 26% to 97%, with a mean of 55.93% (SD = 14.79%).

A. Blind Compliance and PPV

Automation PPV was found to be a significant predictor of blind compliance $t(59) = 9.75, p < .001$. The random-intercept model was used and $\beta_1 = 0.82$. A scatter plot of the extracted data is shown in Figure 2.

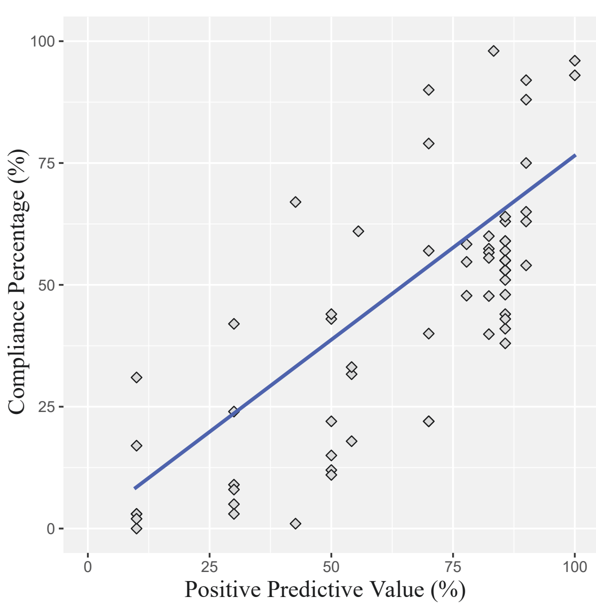


Fig. 2. Operators' blind compliance (Y axis) was plotted against automation's positive predictive value (X axis).

B. Blind Reliance and NPV

Automation NPV was found to be a significant predictor of blind reliance $t(58) = 7.99, p < .001$. The random-intercept model was used and $\beta_1 = 0.89$. A scatter plot of the extracted data is shown in Figure 3.

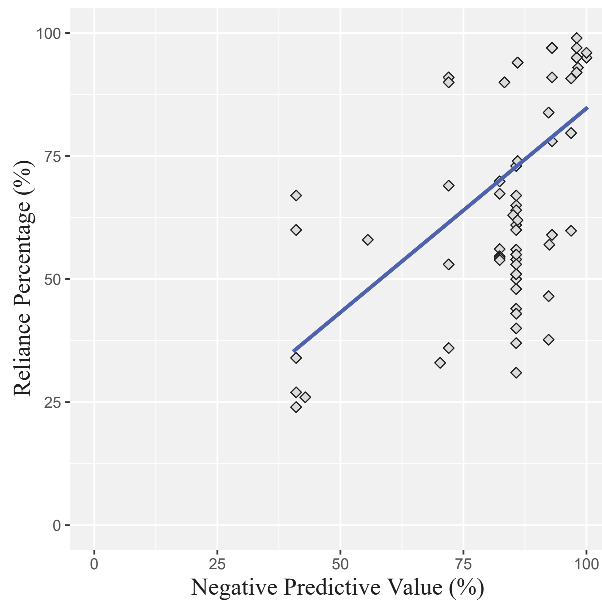


Fig. 3. Operators' blind reliance (Y axis) was plotted against automation's negative predictive value (X axis).

C. Dependence and OSL

Automation OSL was found to be a significant predictor of dependence $t(57) = 3.08, p < .01$. The random-intercept model was used and $\beta_1 = 0.80$. A scatter plot of the extracted data is shown in Figure 4.

IV. Discussion

The objective of this study was to explore how operators' blind behaviors – i.e, blind compliance, blind reliance, and blind dependence – were influenced when using diagnostic automation in dual-task scenarios. We contrasted human operators' observed behaviors against the automation's PPV, NPV, and OSL. Our results show that as PPV, NPV, and OSL increase, human operators' blind compliance, reliance, and dependence increase accordingly. So as automation performs better, the operator is likely to blindly comply with or blindly rely on it more often. This is consistent with previous literature, showing that human operators were adjusting their behavior based on how well the automation was performing [2]. The results are also consistent with the 'probability matching' heuristic [24].

The slope for the relationship between OSL and dependence ($\beta_1 = 0.80$) was lower than the slope for the blind compliance against PPV ($\beta_1 = 0.82$) and the slope for the blind reliance against NPV ($\beta_1 = 0.89$). This result is rather surprising, and could have been because that the dependence behavior is a pooled quantity.

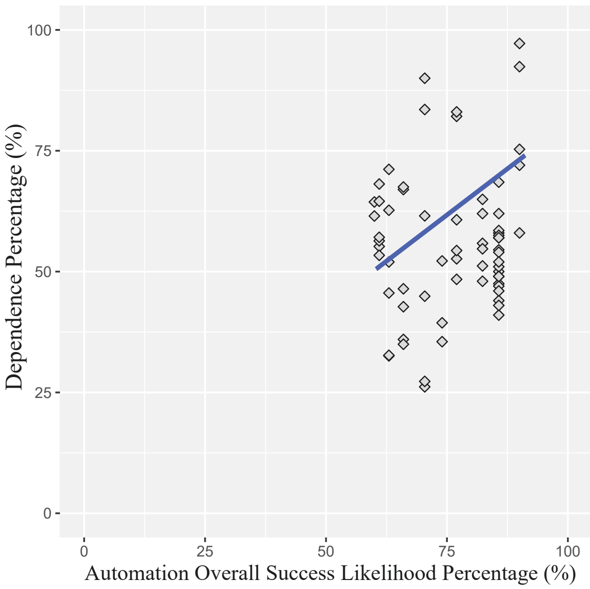


Fig. 4. Operators' blind dependence (Y axis) was plotted against automation's overall success likelihood (X axis).

We recommend that researchers report details of the SDT table (e.g., # of hits, misses, false alarms, and correct rejections) rather than only the overall success likelihood.

This study has the following limitations: First, we focused on blind behaviors, when human operators followed the automation's suggestion and chose not to cross-check with raw information. Second, independent variables other than automation performance levels (i.e., PPV, NPV, and OSL) were treated as noise in our study. Finally, we only included studies that reported full information about the automation's hits, misses, false alarms, and correct rejections.

V. Conclusion

We examined how humans interact with an imperfect automated aid. People adjusted their blind compliance and reliance behaviors as automation became more capable. When examining blind dependence, we found that the operator did not adjust their dependence behavior as much compared to either blind compliance or blind reliance behaviors. We recommend that researchers report the number of automation errors and types (false alarms or misses) rather than automation OSL alone.

Note: References marked with asterisks were used for meta-analysis.

References

[1] Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. Human factors, 39(2), 230-253.

[2] Wickens, C. D., & Dixon, S. R. (2007). The benefits of imperfect diagnostic automation: A synthesis of the literature. Theoretical Issues in Ergonomics Science, 8(3), 201-212.

[3] Yang, X. J., Unhelkar, V. V., Li, K., & Shah, J. A. (2017, March). Evaluating effects of user experience and system transparency on trust in automation. In 2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI (pp. 408-416). IEEE.

[4] Yang, X. J., Schemanske, C., & Searle, C. (2021). Toward Quantifying Trust Dynamics: How People Adjust Their Trust After Moment-to-Moment Interaction With Automation. Human Factors, 0(0). 00187208211034716

[5] 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.

[6] Meyer, J. (2001). Effects of warning validity and proximity on responses to warnings. Human factors, 43(4), 563-572.

[7] Meyer, J. (2004). Conceptual issues in the study of dynamic hazard warnings. Human factors, 46(2), 196-204.

[8] Macmillan, N. A., & Creelman, C. D. (2004). Detection theory: A user's guide. Psychology press.

[9] Tanner Jr, W. P., & Swets, J. A. (1954). A decision-making theory of visual detection. Psychological review, 61(6), 401.

[10] Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group*. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Annals of internal medicine, 151(4), 264-269.

[11] Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. Human factors, 53(5), 517-527.

[12] Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. Human factors, 57(3), 407-434.

[13] Hancock, P. A., Kessler, T. T., Kaplan, A. D., Brill, J. C., & Szalma, J. L. (2021). Evolving trust in robots: specification through sequential and comparative meta-analyses. Human factors, 63(7), 1196-1229.

[14] Kaplan, A. D., Kessler, T. T., Brill, J. C., & Hancock, P. A. (2021). Trust in artificial intelligence: Meta-analytic findings. Human Factors, 00187208211013988.

[15] Hofmann, D. A., Griffin, M. A., & Gavin, M. B. (2000). The application of hierarchical linear modeling to organizational research.

[16] *Barg-Walkow, L. H. (2013). Understanding the role of expectations on human responses to an automated system (Doctoral dissertation, Georgia Institute of Technology).

[17] *Barg-Walkow, L. H., & Rogers, W. A. (2016). The effect of incorrect reliability information on expectations, perceptions, and use of automation. Human factors, 58(2), 242-260.

[18] *Chancey, E. T., Bliss, J. P., Proaps, A. B., & Madhavan, P. (2015). The role of trust as a mediator between system characteristics and response behaviors. Human factors, 57(6), 947-958.

[19] *Du, N., Huang, K. Y., & Yang, X. J. (2020). Not all information is equal: effects of disclosing different types of likelihood information on trust, compliance and reliance, and task performance in human-automation teaming. Human factors, 62(6), 987-1001.

[20] *Gérard, N., & Manzey, D. (2010). Are false alarms not as bad as supposed after all? A study investigating operators' responses to imperfect alarms. In D. de Waard, A. Axelsson, M. Berglund, B. Peters, & C. Weikert (Eds.), Human factors: a system view of human, technology and organisation (pp. 55-69). Maastricht, the Netherlands: Shaker Publishing.

[21] *Manzey, D., Gérard, N., & Wiczorek, R. (2014). Decision-making and response strategies in interaction with alarms: the impact of alarm reliability, availability of alarm validity information and workload. Ergonomics, 57(12), 1833-1855.

[22] *McBride, S. E., Rogers, W. A., & Fisk, A. D. (2011). Understanding the effect of workload on automation use for younger and older adults. Human factors, 53(6), 672-686.

[23] *Wiczorek, R., & Manzey, D. (2014). Supporting attention allocation in multitask environments: Effects of likelihood alarm systems on trust, behavior, and performance. Human factors, 56(7), 1209-1221.

[24] Bliss, J. P., Gilson, R. D., & Deaton, J. E. (1995). Human probability matching behaviour in response to alarms of varying reliability. Ergonomics, 38(11), 2300-2312.

Fig. 5. Data extracted from records [19], [22], [23], [17], [18]. Note: Data calculation is done based on the definitions specified in this paper. They could be different from those originally reported.

Title	Datapoint	Group	Hits	Misses	FAs	CRs	Total Trials	PPV	NPV	OSL	Compliance	Reliance	Dependence	Table/Figure
Du, N., Huang, K. Y., & Yang, X. J. (2020) Not All Information Is Equal: Effects of Disclosing Different Types of Likelihood Information on Trust, Compliance and Reliance, and Task Performance in Human-Automation Teaming	1	Overall Success Likelihood & Low Reliability	13	2	11	24	50	54.2	92.3	74.0	31.7	46.5	39.4	Data is from Table 2.
	2	Overall Success Likelihood & High Reliability	14	1	4	31	50	77.8	96.9	90.0	58.3	79.7	72.0	
	3	Predictive Values & Low Reliability	13	2	11	24	50	54.2	92.3	74.0	17.9	83.9	52.2	
	4	Predicted Values & High Reliability	14	1	4	31	50	77.8	96.9	90.0	47.8	90.8	75.3	
	5	Hit and CR Rates & Low Reliability	13	2	11	24	50	54.2	92.3	74.0	33.1	37.7	35.5	
	6	Hit and CR Rates & High Reliability	14	1	4	31	50	77.8	96.9	90.0	54.7	59.8	58.0	
McBride, S. E., Rogers, W. A., & Fisk, A. D. (2011) Understanding the Effect of Workload on Automation Use for Younger and Older Adults	1	Younger & Low Workload	14	3	3	14	34	82.4	82.4	82.4	39.9	56.1	48.0	Data for Experiment 1 was extracted from Figure 4.
	2	Younger & Moderate Workload	14	3	3	14	34	82.4	82.4	82.4	47.7	54.6	51.2	
	3	Younger & High Workload	14	3	3	14	34	82.4	82.4	82.4	57.4	54.3	55.9	
	4	Older &Low Workload	14	3	3	14	34	82.4	82.4	82.4	56.6	67.4	62.0	Data for Experiment 2 was extracted from Figure 7.
	5	Older & Moderate Workload	14	3	3	14	34	82.4	82.4	82.4	55.5	53.9	54.7	
	6	Older & High Workload	14	3	3	14	34	82.4	82.4	82.4	60.0	69.9	64.9	
Wiczorek, R., & Manzey, D. (2014) Supporting attention allocation in multitask environments: Effects of likelihood alarm systems on trust, behavior, and performance	1	Binary Alert System compliance without cross checking option	29	1	39	31	100	42.6	96.9	60.0	67.0			Data from paragraph on Page 8, under the "Response Rates"
	2	Binary Alert System compliance with cross checking option	29	1	39	31	100	42.6	96.9	60.0	1.0			
Barg-Walkow, L. H., & Rogers, W. A. (2016) The Effect of Incorrect Reliability Information on Expectations, Perceptions, and Use of Automation. We only used Experiment 2, as Experiment 1 was reported in Barg-Wilkow (2013) thesis.	1	Experiment 2 & lower-than & Day 1	120	20	20	120	280	85.7	85.7	85.7	41.0	54.0	47.5	We present the findings from Experiment 1 in the Barg-Wilkow 2013 Thesis. So we extract the compliance from this document in Figure 9b (PDF page 13/19) and reliance in Figure 10b (PDF page 14/19).
	2	Experiment 2 & lower-than & Day 2	120	20	20	120	280	85.7	85.7	85.7	48.0	61.0	54.5	
	3	Experiment 2 & lower-than & Day 3	120	20	20	120	280	85.7	85.7	85.7	51.0	65.0	58.0	
	4	Experiment 2 & lower-than & Day 4	240	40	40	240	560	85.7	85.7	85.7	53.0	64.0	58.5	
	5	Experiment 2 & higher-than & Day 1	240	40	40	240	560	85.7	85.7	85.7	63.0	51.0	57.0	
	6	Experiment 2 & higher-than & Day 2	240	40	40	240	560	85.7	85.7	85.7	55.0	60.0	57.5	
	7	Experiment 2 & higher-than & Day 3	240	40	40	240	560	85.7	85.7	85.7	57.0	67.0	62.0	
	8	Experiment 2 & higher-than & Day 4	240	40	40	240	560	85.7	85.7	85.7	64.0	73.0	68.5	
The role of trust as a mediator between system characteristics and response behaviors Chancey, E. T., Bliss, J. P., Proaps, A. B., & Madhavan, P. (2015) Conference proceedings paper: False Alarms vs. Misses: Subjective Trust as a Mediator between Reliability and Alarm Reaction Measures Chancey, E. T., Bliss, J. P., Proaps, A. B., Liechty, M. & Proaps, A. B. (2015)	1	False Alarm Prone & 60% Reliability	30	0	24	6	60	55.6	100.0	60.0	61.0	95.0	64.4	Please refer to the comment in cells A24:A27. The compliance and reliance data are found in the conference paper on Figures 2 and 3.
	2	False Alarm Prone & 90% Reliability	30	0	6	24	60	83.3	100.0	90.0	98.0	96.0	97.2	
	3	Miss Prone & 60% Reliability	6	24	0	30	60	100.0	55.6	60.0	93.0	58.0	61.5	
	4	Miss Prone & 90% Reliability	24	6	0	30	60	100.0	83.3	90.0	96.0	90.0	92.4	

Fig. 6. Data extracted from record [16], [20], [18]. Note: Data calculation is done based on the definitions specified in this paper. They could be different from those originally reported.

Title	Datapoint	Group	Hits	Misses	FAs	CRs	Total Trials	PPV	NPV	OSL	Compliance	Reliance	Dependence	Table/Figure
Understanding the Role of Expectations on Human Responses to an Automated System. Barg-Wilkow (2013)	1	Explicit Statement & Perceived Low Workload & Day 1	120	20	20	120	280	85.7	85.7	85.7	53.0	37.0	44.0	These compliance data were extracted from Figure 25, on the PDF page 56/98. The reliance data were extracted from Figure 18 on PDF page 58/98.
	2	Explicit Statement & Perceived Moderate Workload & Day 1	120	20	20	120	280	85.7	85.7	85.7	55.0	40.0	47.0	
	3	Explicit Statement & Perceived High Workload & Day 1	120	20	20	120	280	85.7	85.7	85.7	53.0	48.0	50.0	
	4	Initial Exposure & Perceived Low Workload & Day 1	120	20	20	120	280	85.7	85.7	85.7	53.0	31.0	41.0	
	5	Initial Exposure & Perceived Moderate Workload & Day 1	120	20	20	120	280	85.7	85.7	85.7	38.0	44.0	43.0	
	6	Initial Exposure & Perceived High Workload & Day 1	120	20	20	120	280	85.7	85.7	85.7	44.0	56.0	51.0	
	7	Explicit Statement & Perceived Low Workload & Day 2	120	20	20	120	280	85.7	85.7	85.7	59.0	50.0	54.0	
	8	Explicit Statement & Perrceived Moderate Workload & Day 2	120	20	20	120	280	85.7	85.7	85.7	55.0	43.0	49.0	
	9	Explicit Statement & Perceived High Workload & Day 2	120	20	20	120	280	85.7	85.7	85.7	59.0	55.0	57.0	
	10	Initial Exposure & Perceived Low Workload & Day 2	120	20	20	120	280	85.7	85.7	85.7	53.0	51.0	52.0	
	11	Initial Exposure & Perceived Moderate Workload & Day 2	120	20	20	120	280	85.7	85.7	85.7	43.0	53.0	46.0	
	12	Initial Exposure & Perceived High Workload & Day 2	120	20	20	120	280	85.7	85.7	85.7	55.0	43.0	49.0	
Are false alarms not as bad as supposed after all? A study investigating operators' responses to imperfect alarms Gérard, N., & Manzey, D. (2010)	1	Base Rate .05	4	1	38	57	100	10.0	98.3	61.0	3.0	93.0	55.2	Figures 2 and 3 were used for compliance and reliance
	2	Base Rate .18	14	4	33	49	100	30.0	92.5	63.0	5.0	57.0	32.6	
	3	Base Rate .33	26	7	27	40	100	50.0	85.1	66.0	12.0	63.0	36.0	
	4	Base Rate .54	43	11	18	26	98	70.0	70.3	70.4	22.0	33.0	26.2	
	5	Base Rate .81	65	16	7	12	100	90.0	42.9	77.0	63.0	26.0	52.6	
Decision-making and response strategies in interaction with alarms: the impact of alarm reliability, availability of alarm validity information and workload. Manzey, D., Gérard, N., & Wiczorek, R. (2014)	1	Exp 1 Block 1	4	1	38	57	100	10.0	98.0	61.0	31.0	95.0	68.1	Data from Figure 2 provided compliance and reliance for Exp 1, Block 1 and 2.
	2	Exp 1 Block 1	14	4	33	49	100	30.0	93.0	63.0	42.0	97.0	71.2	
	3	Exp 1 Block 1	26	7	27	40	100	50.0	86.0	66.0	43.0	94.0	67.0	
	4	Exp 1 Block 1	43	11	18	26	98	70.0	72.0	70.4	79.0	91.0	83.5	
	5	Exp 1 Block 1	65	16	7	12	100	90.0	41.0	77.0	88.0	67.0	82.1	
	6	Exp 1 Block 2	4	1	38	57	100	10.0	98.0	61.0	17.0	99.0	64.6	
	7	Exp 1 Block 2	14	4	33	49	100	30.0	93.0	63.0	24.0	97.0	62.7	
	8	Exp 1 Block 2	26	7	27	40	100	50.0	86.0	66.0	44.0	94.0	67.5	
	9	Exp 1 Block 2	43	11	18	26	98	70.0	72.0	70.4	90.0	90.0	90.0	
	10	Exp 1 Block 2	65	16	7	12	100	90.0	41.0	77.0	92.0	60.0	83.0	
	11	Exp 2 blind behaviors	4	1	38	57	100	10.0	98.0	61.0	0.0	92.0	53.4	Data from Figure 4 (PDF page 14/24) has the compliance and reliance for Exp 2. We are using the "Direct Response" during alarm for compliance and "No Response" during non-alarm for reliance.
	12	Exp 2 blind behaviors	14	4	33	49	100	30.0	93.0	63.0	3.0	59.0	32.7	
	13	Exp 2 blind behaviors	26	7	27	40	100	50.0	86.0	66.0	11.0	62.0	35.0	
	14	Exp 2 blind behaviors	43	11	18	26	98	70.0	72.0	70.4	22.0	36.0	27.3	
	15	Exp 2 blind behaviors	65	16	7	12	100	90.0	41.0	77.0	65.0	27.0	54.4	
	16	Exp 3 Blind Behaviors	4	1	38	57	100	10.0	98.0	61.0	3.0	95.0	56.4	Data from Figure 5 (PDF page 17/24) has the compliance and reliance for Exp 3. We are using the "Direct Response" during alarm for compliance and "No Response" during non-alarm for reliance.
	17	Exp 3 Blind Behaviors	14	4	33	49	100	30.0	93.0	63.0	9.0	78.0	45.6	
	18	Exp 3 Blind Behaviors	26	7	27	40	100	50.0	86.0	66.0	15.0	74.0	42.7	
	19	Exp 3 Blind Behaviors	43	11	18	26	98	70.0	72.0	70.4	40.0	53.0	44.9	
	20	Exp 3 Blind Behaviors	65	16	7	12	100	90.0	41.0	77.0	75.0	24.0	60.7	
	21	Exp 4 Blind Behaviors	4	1	38	57	100	10.0	98.0	61.0	2.0	97.0	57.1	Data from Figure 6 (PDF page 20/24) has the compliance and reliance for Exp 4. We are using the "Direct Response" during alarm for compliance and "No Response" during non-alarm for reliance.
	22	Exp 4 Blind Behaviors	14	4	33	49	100	30.0	93.0	63.0	8.0	91.0	52.0	
	23	Exp 4 Blind Behaviors	26	7	27	40	100	50.0	86.0	66.0	22.0	74.0	46.4	
	24	Exp 4 Blind Behaviors	43	11	18	26	98	70.0	72.0	70.4	57.0	69.0	61.5	
	25	Exp 4 Blind Behaviors	65	16	7	12	100	90.0	41.0	77.0	54.0	34.0	48.4	

6-8 v3.pdf