**MEDWIN PUBLISHERS**
Committed to Create Value for Researchers

# An Influence Model of the Human-Automation Team: Effects of Workload and Automation Reliability, Transparency and Degree

**Wickens CD[1]\*, Sargent R[2] and Walters B[2]**

[1]Psychology Department, Colorado State University, USA

[2]Huntington Ingalls Industry, USA

**\*Corresponding author:** Christopher D Wickens, Psychology Department, Colorado State University, USA, Tel: 01-303-775-5144; Email: pandawickens94@aol.com

## Abstract

Four important variables that influence performance of a human-automation team (HAT) are the *workload (resource) demands* of the task and environment, and the *reliability*, *degree*, and *transparency* of the automation that assists the human. In this paper, first, we present a model that predicts how these variables influence the human contribution to performance of the HAT, as mediated by changes in trust, dependence, and situation awareness (SA) and experienced workload (dependent variables). These factors vary in their strength of influence and their interaction with each other. Then we describe in greater depth how a meta- analysis of 50 studies has revealed differences in the strength of influence of automation transparency on those dependent variables. In particular, transparency has a large effect on improving the accuracy of performance, increasing trust and situation awareness. Transparency has different benefits for performance in routine situations, where it improves accuracy of the human-automation team, than for situations when automation unexpectedly fails, where it decreases the time for failure recovery. Finally, we illustrate how the model accommodates differences in discrimination task difficulty revealed in an experiment on decision aiding for nautical collision avoidance. This shows the benefits of the model in predicting the tradeoff of factors influencing human-automation team performance.

**Keywords:** Human-automation; Automation *Transparency*; Reliability; *Reliance* or *Compliance*

**Abbreviations:** DOA: Degree of Automation; PTC: Positive Train Control; HAI: Human-Automaton Interaction; AI: Artificial Intelligence; SA: Situation Awareness; ATP: Automation Transparency.

## Introduction

Human-automaton interaction (HAI) has been the subject of a myriad of articles over the past half century, beginning with a unique focus on aircraft automation [1] soon addressing issues in health care [2] industry [3] driving [4] security, business or legal decisions, education [5] consumer products [6] air traffic control [7] and the military

[8]. Most recently, an explicit focus on the capabilities of artificial intelligence (AI) algorithms and machine learning have amplified this interest [9,10]. The focus on AI has been extended to applications such as Chat GPT.

Although there are many reasons for the choice – to automate or not, and if to automate, to what degree or at what level – ultimately one primary reason for the choice is the performance of the human-automation team (HAT), as typically assessed by measures of accuracy and/or speed. Among the thousands of articles on the topic, a large number have examined the influence of certain properties of automation on measures of performance of the HAT: for

An Influence Model of the Human-Automation Team: Effects of Workload and Automation Reliability, Transparency and Degree

Ergonomics Int J

example, the influence of *reliability* [11-13] increasing which is found to generally improve performance; or the influence of *automation transparency*, also generally found to produce a benefit [14,15]. Typically, these effects are reported with statistical measures of "significance". We argue here that what is needed as a next step is to turn these statistics into quantitative estimates of the *degree of benefit* (or cost) of specific automation features. These estimates can be offered at the micro-level: for example, implementing a certain form of transparency can shorten a driver's take-over time with a self-driving car by XXX msec; or alternatively at a macro-level: for example, incorporating positive train control (PTC) in the locomotive cab, can reduce the frequency of derailment accidents by YY%. Such numbers can then be traded off against the financial costs of implementing automation features, to assess, from an actuarial standpoint, the net expected "value" (or cost) of adopting the feature or type of automation in question.

A complication to this kind of quantification arises because the various features of automation do not act in isolation from each other. Instead, they interact in sometimes unpredictable ways. For example, adding an autopilot steering control to a car may produce large benefits if it is highly reliable, but diminishing benefits if it is not, compared to adding auto-speed control. Hence, any quantitative predictive model should account for these linkages or interactions.

In the following, we present the *framework* for such a model. We emphasize framework, because what we present is far from complete and fully validated in terms of providing quantitative estimates of the strength with which a particular feature of automation will influence metrics of performance. This is particularly true when it comes to interactions between influences. Instead, in the following we:

- Highlight what we view are some of the most important interactions.
- Provide some rough estimates of relative weights for certain factors (and invite others to do so).
- Incorporate important intervening cognitive constructs between automation features and performance, in particular trust, dependence, situation awareness (SA) and workload.
- Address a subset of three features that we consider to be among the most important features of automation that influence HAT performance.

After we describe this influence model in some detail, we then report new data we have acquired on the specific quantitative, influences of transparency and automation reliability, and the implications of these data for the model.

## The Hat Influence Model

The model of the influence on HAT performance of three properties of automation, coupled with the workload of the task and environmental is shown in Figure 1. To the far left we depict the effects of the difficulty (resource demands) of the task which is to be automated, along with the difficulty (cognitive demands or workload) imposed by the overall environment in which the task is performed. This latter factor can include the demands of concurrent tasks imposed while performing or assisting with the automated task.
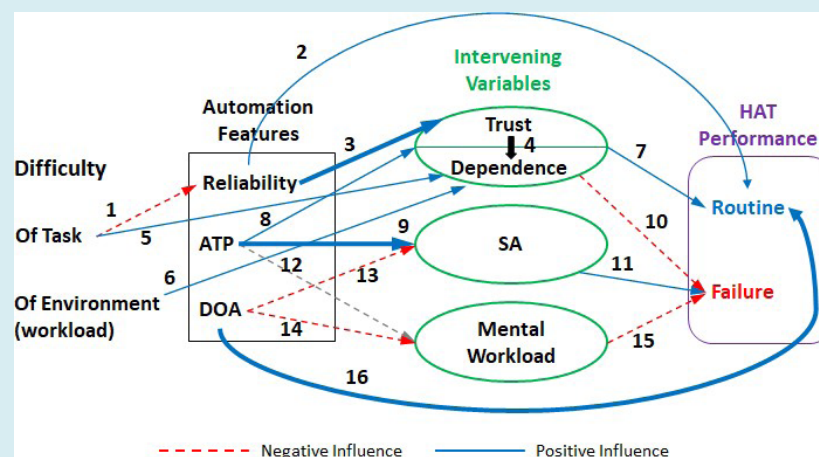


**Figure 1:** The Influence Model of HAT Performance.
**Note:** The differences in the widths of the links indicate the strength of inferred effect, to be discussed later in the paper.

In the left box, we represent the three key interrelated features of the automated system, its reliability, the transparency imposed on its functioning (ATP) and the degree of automation (DOA), a variable that is increased both by the number of stages of human information processing automation is designed to assist, and by the level

Wickens CD, et al. An Influence Model of the Human-Automation Team: Effects of Workload and Automation Reliability, Transparency and Degree. Ergonomics Int J 2023, 7(4): 000312.

Copyright© Wickens CD, et al.

of automation responsibility within each and all stages and the level of automation authority within each stage [16-18]. The green ovals in the center represent four intervening dependent variables, influenced by those properties of the automation, task and the environment. Of these, trust and dependence are grouped closely together because these two constructs, one purely cognitive and the other inferred from behavior, are often closely linked. SA and mental workload are constructs that have long been linked in the study of the HAT [19]. In the box to the right are two different aspects of the performance of the HAT (as typically measured by speed and/or accuracy).

*Routine performance* is that which the automation was designed for; it is often thought of as a measure of productivity, such as the time to travel to a destination with a self-driving car, or the accuracy of classifying a threat in the battlefield. *Failure performance* refers to the fluency with which the human can respond to a (typically) unexpected "failure" of the automation (including a failure of the power supply supporting its functions).

Between these elements of the HAT representation are causal links that depict inferred relationships between an increase in the commodity to the left and the level of the commodity to the right. A blue solid arrow indicates a relationship assumed to be positive (an increase in the cause leads to an increase in the effect); and a red dashed arrow connects elements with an inverse relationship (increase → decrease). Many of these 16 causal links are intuitive; and many are supported by research examples provided below. However, the *relative* strengths of these are often not well established, an important issue when actual performance predictions are to be made. We describe and number these links in the following text, generally moving down the figure from top to bottom and left-to-right. Inverse links are indicted by red arrows in the text. The difference in link width (indicating inferred inference strength) will be discussed at the end of the paper, after the empirical data are presented.

**Link 1:** Task difficulty → Automation reliability. This link is highly intuitive. The more challenging the task is, the less likely automation will be able to perform it perfectly. In target classification, more similar features between candidates (e.g., faces) will create greater challenges for machine-vision software.

**Link 2:** Reliability of automation → Routine performance. More reliable automation may result from increased algorithm sophistication or increased machine learning classification rate. This arrow is curved, bypassing the human component. Whatever the source of better (more reliable) automation performance, this curved link indicates that it is expected to yield better performance in the routine task for which it was designed.

**Link 3:** Reliability → Trust. It is both well established [20]

and intuitive that more reliable automation will be provided a higher subjective rating of trust by the human.

**Link 4:** Trust → Dependence. Subjective trust and objectively measured *dependence* (often defined by the terms *reliance* or *compliance*; [21] tend to be closely linked, but are sometimes conflated in the literature. Indeed, they are distinctly different intervening variables, and often dissociate [22]. This is reflected by the different causal variables in the figure that are inputs to each, in the two links we describe next.

**Link 5:** Difficulty (workload) of the task → Automation dependence. It is both intuitive and reflected by research [23,24] that more difficult discrimination tasks yield greater dependence on automation, even as that automation may not be judged to be more reliable.

**Link 6:** Difficulty (workload) of the environment → Automation dependence. This link bypasses "trust" and flows directly to dependence, reflecting the fact that in high demand environments, operators may be forced to depend on (use) automation that they do not fully trust.

**Link 7:** Dependence → Routine performance. The strength of this link depends, in part, upon the disparity in performance between automation alone and the human alone. The greater is the former relative to the latter, the stronger the typical link is found to be [23-25]. However, this moderator does not imply that the human will accurately calibrate dependence to automation reliability. Operators often falling well short of that optimal dependence, such that a greater gap in performance between the two agents often a produces a greater shortfall between optimal and obtained HAT performance [11,23,24].

**Link 8:** Automation transparency (ATP) Trust. → This link is intuitive: Generally, people "like" automation that can explain how it is operating, and why, for example, automation arrived at its particular judgment or decision. As evident from link 4, the increased trust typically produces increased dependence.

**Link 9**: Transparency → SA. This link is almost a sine qua non: providing more information about what automation is doing will directly improve SA so long as that information is attended and understood. And providing more information about how automation works, should make the changes that automation is implementing more interpretable. Both "how it works" and "what it is doing" are vital components contributing to SA. [26].

**Link 10:** Dependence → Failure performance. It is intuitive that the more dependent one is on automation (as reflected, for example by greater investment of resources into concurrent tasks), the less vigilant one is of monitoring the concurrent automation. Also, the more dependent the user is upon its proper functioning, the less effective the user will be in intervening when automation fails [27].

**Link 11:** Situation awareness → Failure performance. This link, well validated by literature, and represented in the tradeoff lumberjack model [28] is based on the thinking that when things go wrong, and the operator must jump back

into the loop, this will be more effective and fluent if the operator is aware of what automation is doing at the time of the failure. Notice that there is no link from SA to routine performance. SA is a more important construct when things go wrong, than when automation supports their normal HAT functioning [16].

**Link 12:** ATP → Workload. The polarity (increase or decrease) of this link is uncertain because of the fact that transparency, while presumably alleviating the workload of maintaining SA might also, paradoxically increase the workload required to process the added displayed transparency information, hence offsetting any benefit to workload reduction [29].

**Link 13:** DOA → SA. This link is negative or inhibitory. A higher DOA "does more work" and hence, presumably the human does less. Doing less cognitive work in the HAT can lead to a reduction in SA because of the "generation effect" [30]: one tends to be less aware of the state of a dynamic situation when one is not generating actions pertaining to that system (and is simply monitoring automation performing those same actions).

**Link 14:** DOA→ Mental workload. The reason for this inhibitory relationship was stated in the previous link between DOA and SA.

**Link 15:** Mental workload → Failure response. Here we argue that fluent intervention when things go wrong typically requires cognitive resources. Any characteristic that produces more mental load and inhibits that "reserve capacity" available to diagnose a failure (higher workload) will degrade failure performance.

**Link 16:** DOA → Routine performance. This link, like curved Link 2 above, directly bypasses the human and is based on the assumption that, of the several reasons to invoke automation (or "more" automation: a higher DOA), one of the most prominent is to improve the "productivity" for which the automation was intended.

### New Data

The links in Figure 1 are an important first step in defining this HAT influence model. However, the representation in the figure says nothing about two important and related issues that define the requirements for a computational model. First, there is no current representation of influence strength in any of the links, an element that is important if one is to make informed decisions regarding the overall utility of incorporating an automation feature (e.g., ATP, or increasing the DOA). Second, without representation of influence strengths, when two influences conflict, there is no way to determine "who wins". As a specific example, consider the effect of increasing automation task difficulty. On the one hand, this is likely to decrease the reliability of automation and hence decrease user trust and dependence (links 2, 3 and 4). On the other hand, a more difficult task will directly degrade the capability of the human to perform the same

task manually, and hence likely to increase dependence upon automation (link 5), even as it may not be fully trusted.

Below, we present new data, first regarding the strengths of transparency benefit links and then regarding the trade-off effects of discrimination task difficulty.

### Transparency Meta-Analysis

A meta-analysis was completed using the results of 50 studies that have examined the benefits (or lack thereof) of incorporating transparency into the HAT. Details are reported in [26]. From the 81 effect sizes that were derived (some studies reported more than one effect), we were able to report the effect sizes, representing ATP effect strength on measures of the four intervening variables shown in Figure 1 as well as performance impacts (error rate, ER, and response times, RT) on HAT task performance. These data are shown in Table 1.

| Dependent Variable | Mean Effect Size (Cohen's d) | N | SD | SE | 2SE |
|---|---|---|---|---|---|
| ER | -0.96 | 32 | 2.5 | 0.44 | 0.88 |
| RT | -0.53 | 48 | 1.22 | 0.18 | 0.35 |
| Trust | 0.79 | 43 | 0.89 | 0.14 | 0.27 |
| Dependence | 0.45 | 46 | 1.25 | 0.18 | 0.37 |
| Workload | -0.15 | 28 | 1.25 | 0.24 | 0.47 |
| SA | 1.06 | 18 | 1.57 | 0.37 | 0.74 |

**Table 1:** Effect Sizes of Either Imposing Transparency or Increasing the Degree of Transparency.
p<.05, p<.01

The table clearly reveals the large effect size of transparency on reducing error and increasing SA, the medium effect size on decreasing RT and increasing both trust and dependence, and essentially no effect on workload. These differential influences are reflected in the width of the links shown in Figure 1. Some additional differences in link strength in the figure are derived from the meta-analysis on DOA reported by Onnasch, et al [28]. The influence model shown in Figure 1 also makes the important distinction between routine performance and failure performance; and more details of the meta-analysis results reveals why this distinction is important. When performance of the routine task is examined (links 8,4,7) incorporating transparency has a large effect on reducing error rate ($d$ = -1.02) but only a medium, beneficial effect on performance RT (d=-0.41). However, when the effect on failure performance is examined (links 9,11), the benefits to speed and accuracy are reversed: The benefit for RT is large ($d$ = -0.90) and for ER is only medium ($d$ = -0.44). This finding seems particularly

Wickens CD, et al. An Influence Model of the Human-Automation Team: Effects of Workload and Automation Reliability, Transparency and Degree. Ergonomics Int J 2023, 7(4): 000312.

Copyright© Wickens CD, et al.

important in time critical environments [31,32], such as driving with an autopilot [31]. Careful incorporation of transparency can substantially reduce the "takeover time" for an automation failure; and when travelling at a high speed, "seconds matter".

## Opposing Effects of Discrimination Task Difficulty

As noted above, the link chain from task difficulty to automation reliability to trust and dependence (1,3,4) makes opposite predictions to that from difficulty to dependence (5). Which influence component has the stronger effect when the difficulty of a detection or discrimination task is varied, and the raw data for that task is visible to both machine and human?

In a recent unpublished experiment, Pharmer and Wickens asked participants to make judgment of appropriate nautical collision avoidance maneuvers, when they could both see the direct ship trajectories (on a 2D plan view display) and see the advice of an automated decision aid. The collision problems varied in the difficulty of discriminating the better from the less optimal turn direction. Importantly, automation performance (reliability) was also degraded in proportion to that discrimination difficulty (Link 1). Link chain (1,3,4) would predict that dependence on automation should decline as difficulty increases. In contrast, link 5 predicts that it should increase. Here, the data were clear, in support of the dominance of the link chain [1,3,4], over link 5. As the difficulty of discrimination increased and automation reliability hence declined from 100% (easy) to 72% (medium), to 58% (difficult), the dependence on automation (the extent to which the human complied with the automation's imperfect recommendation) declined proportionally and significantly, from 81% to 73% to 64%. This difference is reflected in the thickness of influence weights in Figure 1. It may be interpreted as reflecting humans' overconfidence in their own ability to judge the raw data, in the face of an increasingly difficult judgment task.

## Conclusion

In conclusion, The Influence Model presents a plausible representation of the influence of four key features: the task performed by the HAT and 3 features of the automation member of the HAT, as they influence human cognition and HAT performance. We have confidence of the model's validity in terms of the polarity of influences, but it is less well documented in its current form, in terms of influence strength. The meta-analysis presented on the effects of transparency illustrates how this can be done, and such strengths have also been assessed via meta-analysis for degree of automation [28]. It is important that the current

model be augmented in this regard concerning the influence of task difficulty (here reflected in discrimination difficulty) and, of course that the global predictions of the model be validated through experiments, manipulating combinations of variables.

## Acknowledgments

## References

1. Wiener EL, Curry RE (1980) Flight-deck automation: Promises and problems. Ergonomics 23(10): 995-1011.

2. Morrow D, Wickens CD, North R (2005) Reducing and mitigating human error in medicine. In: Nickerson RS (Ed.), Review of human factors and ergonomics. Human Factors and Ergonomics Society 1.

3. Strobhar D (2012). Human factors in process plant operation. Momentum Press.

4. Fisher DL, Horrey WJ, Lee JD, Regan MA (2020) Handbook of human factors for intelligent and automated vehicles. CRC Press.

5. Conati C, Porayska-Pomsta K, Mavrikis M (2018) AI in Education needs interpretable machine learning: Lessons from Open Learner Modelling. ArXiv: 1807.00154.

6. Nof S (2009) Springer handbook of automation. Springer.

7. Wickens CD, Mavor AS, Parasuraman R, McGee JP (1998) The future of air traffic control: Human operators and automation. National Academy Press.

8. Chen J, Lakhman S, Stowers K, Sellpwotz A, Wright J, et al (2018) Situation awareness based agent transparency and human-autonomy teaming effectiveness. Theoretical Issues in Ergonomic Science 19(3): 259-282.

9. Schneiderman B (2021) Human-Centered AI. Oxford University Press, UK.

10. Rosen P, Heinold E, Fries-Tersch E, Moore P, Wischniewski S (2022) Advanced robotics, artificial intelligence and the automation of tasks. European Agency for Safety and Health at Work.

11. Hutchinson J, Strickland L, Farrell S, Loft S (2022) The perception of automation reliability and acceptance of automated advice. Human Factors.

12. Oakley B, Mouloua M, Hancock P (2003) Effects of automation reliability on human monitoring performance. Proceedings of the human factors and ergonomics society annual meeting 47(1): 188-190.

13. Ross JM, Szalma JL, Hancock PA, Barnett JS, Taylor G (2008) The effect of automation reliability on user automation trust and reliance in a search-and-rescue scenario. Proceedings of the human factors and ergonomics society annual meeting 52(19): 1340-1344.

14. van de Merwe K, Mallam S, Nazir S (2022) Agent Transparency, Situation Awareness, Mental Workload, and Operator Performance: A Systematic Literature Review. Human Factors.

15. Bhaskara A, Skinner M, Loft S (2020) Agent transparency: A review of current theory and evidence. IEEE Transactions on Human-Machine Systems 50(3): 215-224.

16. Parasuraman R, Sheridan TB, Wickens CD (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.

17. Onnasch L, Wickens CD, LiH , Manzey D (2014) Human performance consequences of stages and levels of automation: An integrated meta-analysis. Human Factors 56(3): 476-488.

18. Kaber DB (2018) Issues in human-automation interaction modeling: Presumptive aspects of frameworks of types and levels of automation. Journal of Cognitive Engineering and Decision Making 12(1): 7-24.

19. Parasuraman R, Sheridan TB, Wickens CD (2008) Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. Journal of Cognitive Engineering and Decision Making 2(2): 140-160.

20. Hoff KA, Bashir M (2015) Trust in automation: Integrating empirical evidence on factors that influence trust. Human Factors 57(3): 407-434.

21. Meyer J, Lee JD (2013) Trust, reliance and compliance.

In: Lee JD, et al. (Eds.), The Oxford handbook of cognitive engineering. Oxford University Press, pp: 109-124.

22. Lee JD, See J(2004) Trust in automation and technology: Designing for appropriate reliance. Human Factors 46(1): 50-80.

23. Boskemper MM, Bartlett ML, McCarley JS (2021) Measuring the efficiency of automation-aided performance in a simulated baggage screening task. Human Factors 64(6): 945-961.

24. Bartlett ML, McCarley JS (2020) Ironic efficiency in automation-aided signal detection. Ergonomics 64(1): 1-10.

25. Wickens CD, McCarley JS, Gutzwiller RS (2023) Applied attention theory. 2nd (Edn.), CRC Press, Roultledge taylor & francis groups.

26. Sargent R, Walters B, Wickens CD (2023) Meta-analysis qualifying and quantifying the benefits of automation transparency to enhance models of human performance. Proceedings of HCI International.

27. Endsley MR, Kiris EO (1995) The out-of-the-loop performance problem and level of control in automation. Human factors 37(2): 381-394.

28. Onnasch L, Wickens CD, LiH , Manzey D (2014) Human performance consequences of stages and levels of automation: An integrated meta-analysis. Human Factors 56(3): 476-488.

29. Kunze A, Summerskill SJ, Marshall R, Filtness AJ (2019) Automation transparency: Implications of uncertainty communication for human-automation interaction and interfaces. Ergonomics 62(3): 345-360.

30. Slamecka NJ, Graf P (1978) The generation effect: Delineation of a phenomenon. Journal of Experimental Psychology: Human Learning and Memory 4(6): 592-604.

31. Eriksson A, Stanton NA (2017) Takeover time in highly automated vehicles: Noncritical transitions to and from manual control. Human Factors 59(4): 689-705.

32. Landry S (2021) Human factors in aviation. In: Salvendy G, et al. (Eds.), Handbook of human factors. 5th(Edn.),CRC Press.