Neural Correlates of Human Trust in Automation

Catherine E. Devlin

Advisors: Dr. Susan Davis, University of Dayton Psychology Department
Mr. Justin Estepp, WPAFB Research Laboratory

Research Objective

To find an objective measure of human trust in automation by collecting Electroencephalography (EEG), performance, workload, and subjective data.

Introduction

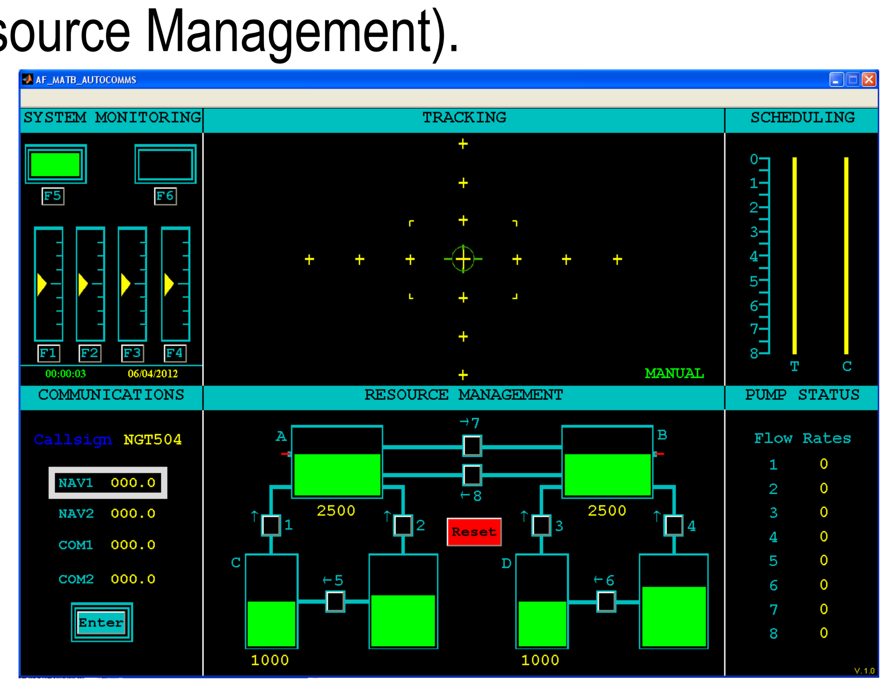
•Previous research exists about interpersonal trust and its neural correlates (Krueger, 2007).
•Very minimal research exists about human trust in automation due to the subjective nature of trust.
•To measure trust in automation, the participant and the automation must develop a calibrated trusting relationship, characterized by vulnerability and uncertainty (Lee and See, 2004) .

Participants

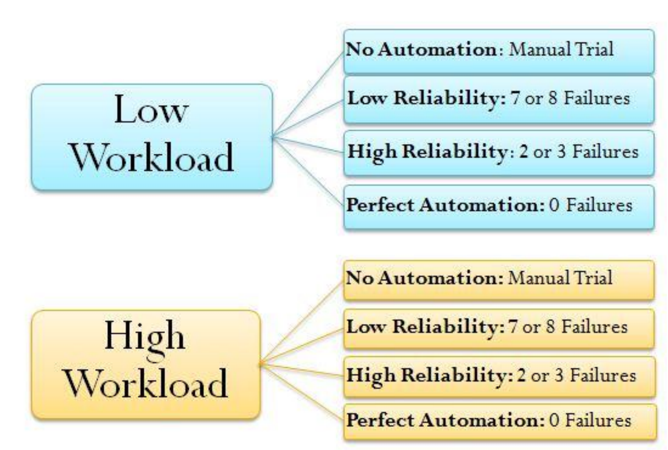
•10 participants completed the study (4 female and 6 male).
•Participants' age ranged from 19-29 with a mean age of 22.7 years old.

Procedure and Materials

•Participants were trained on AF_MATB, a multi-battery attribute (Miller, 2010). AF_MATB consists of a visual task (System Monitoring), an auditory task (Communications), a compensatory task (Tracking), and an executive function task (Resource Management).



•After the participant was trained to a asymptotic performance, they were introduced to the automation. Each participant had time to train and understand the automation.
•The resource management sub-task became an automated task with four varying reliabilities: no automation, low reliability, high reliability, and perfect automation.
•The automation reliabilities correspond to the number of times the automation will fail. A failure is when the resource management fuel tanks deviate from their target values.
•We had a 2x4 factorial design. Low Workload refers to a level of AF_MATB that is easy and High Workload refers to each participant's individual titration point, found through their asymptotic performance scores.
•Further, each participant is given a set of performance thresholds that they must meet. These points are selected based off their performance on previous training days and inform the participant that the automation will help them achieve these thresholds.
•In order to develop a trusting relationship, vulnerability and uncertainty must be present. In order to make the participant feel vulnerable, on the day of data collection we inform them that because of their hard work during training they have earned an endowment of $160 dollars. If they do not meet their thresholds they consequently lose $10, per trial. Uncertainty is present because the automation failures are very unpredictable and can have catastrophic effects.
•The BioSemi Active Two system to collect EEG data from 128 electrodes placed on a cap.
•The participant completed 16 trials, completing two blocks of all conditions. After each trial, the participants answer a series of subjective measures including NASA Task Load Index (TLX) (Hart and Staveland, 1988) and The Trust in Automation Inventory (Jian, Bisantz and Drury, 2000). TLX measures workload, while The Trust in Automation Inventory measures perceived trust.



Results

•*Performance:* we found main effects for workload, automation, and their interaction. Overall, performance goes up with more automation.

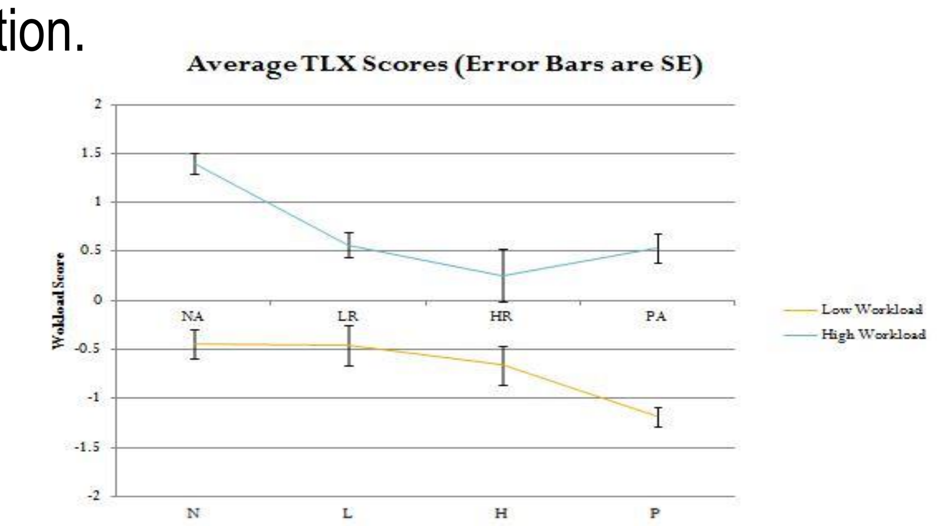
Performance:

Workload:
$F(1,9) = 469.25$,
$p < 0.01, \eta_p^2 = 0.981$

Automation:
$F(3,27) = 20.38$,
$p < 0.01, \eta_p^2 = 0.694$

Interaction:
$F(3,27) = 10.28$,
$p < 0.01, \eta_p^2 = 0.533$

•*Workload (TLX):* As the data shows, there are main effects of workload, automation, and their interaction. Overall, perceived workload decreases with more automation.



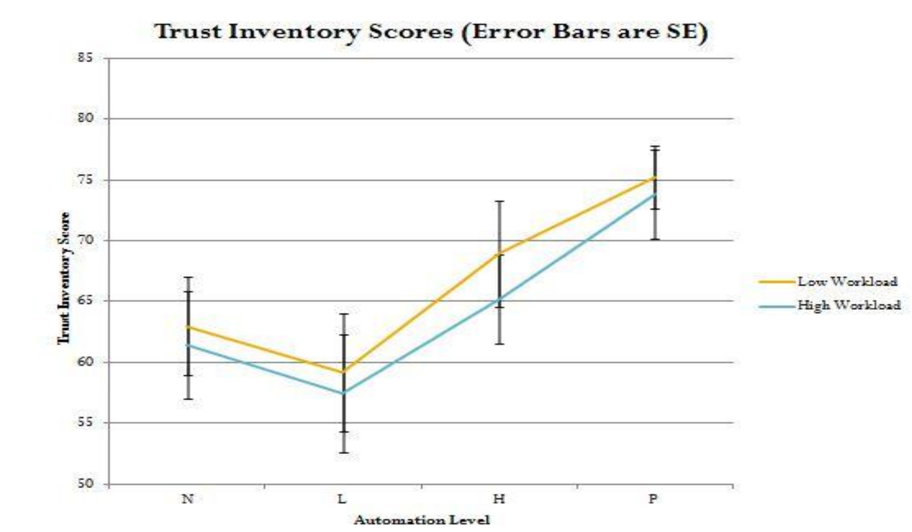
TLX:

Workload:
$F(1,9) = 44.82$,
$p < 0.01, \eta_p^2 = 0.833$

Automation:
$F(3,27) = 10.78$,
$p < 0.01, \eta_p^2 = 0.545$

Interaction:
$F(3,27) = 4.15$,
$p = 0.015, \eta_p^2 = 0.315$

•*Trust Inventory:* there was a marginal effect of workload, a significant main effect of automation, and no significant main effect of their interaction. Perceived trust increases as the automation level increases.



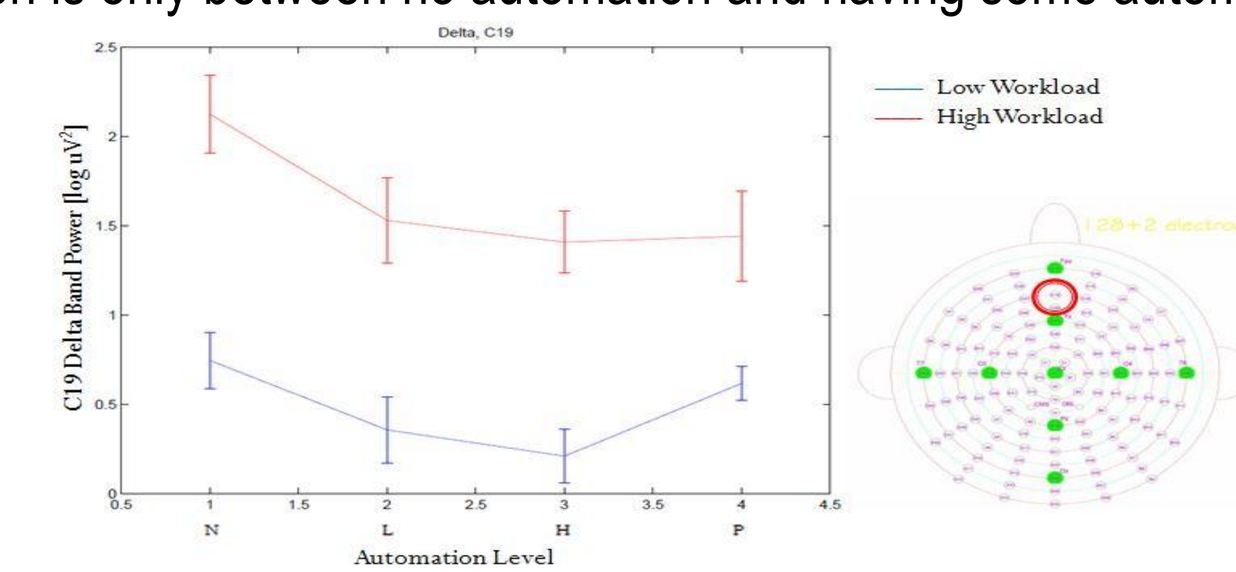
Trust Inventory:

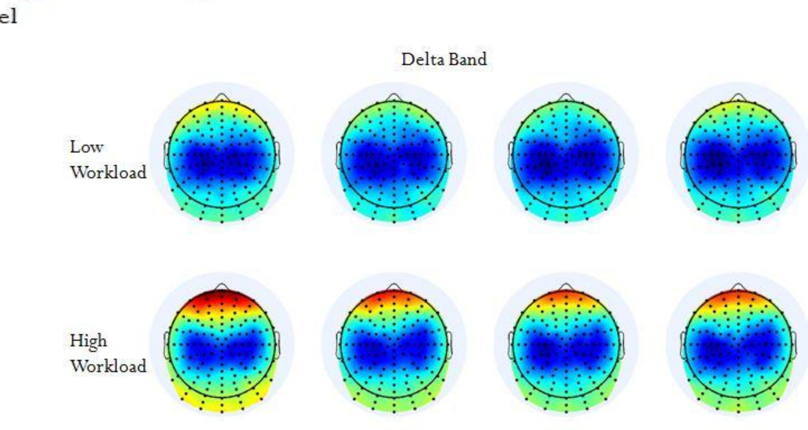
Workload:
$F(1,9) = 5.04$,
$p = .051, \eta_p^2 = 0.359$

Automation:
$F(3,27) = 8.09$,
$p < 0.01, \eta_p^2 = 0.473$

Interaction:
$F(3,27) = .268$,
$p = .848, \eta_p^2 = 0.029$

EEG: The only EEG bands, which correspond to frequency ranges, that had main effects for automation and workload were Delta bands, which correspond to the pre-frontal cortex area. However, the main effect for automation is only between no automation and having some automation, in general.





Discussion

Overall, I believe that I have created a successful protocol for objectively influencing trust through the automation levels. The subjective trust inventory shows us that we have influenced trust by altering the automation level. However, we do not see this same trend in the performance data, TLX, or EEG data. Under the conditions of calibrated trust, I found that automation reduces workload and the changes in EEG activity are a result of workload changes, and not automation levels. This current analysis has only revealed a neurophysiological correlate in EEG data for automation vs. no automation. The correlate for trust in automation (as automation reliability increases) has not yet been established.

References

Hart, S. G. & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.) *Human Mental Workload,* 139-183. Elsevier Science: NorthHolland.

Jian, J., Bisantz,A. and Drury, C. (2000) Foundations for an empirically determined scale of trust in automated systems. Int. J. Cogn. Ergon.,4, 53-71.

Krueger, F., McCabe, K., Moll, J., Kriegskorte, N., Zahn, R., Strenziok, M., et al. (2007). Neural correlates of trust. *Proceedings of the National Academy of Sciences USA, 104,* 20085-20089.

Lee, J.D. & See, K.A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46(1),* 50-80.

Miller, W.D. (2010). The U.S. Air Force-developed adaptation of the Multi-Attribute Task Battery for the assessment of human operator workload and strategic behavior. (Tech. Rep. No. AFRL-RH-WP-TR-2010-0133).