Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 55-41 (2022) 107–112

A Heuristic Strategy for Cognitive State-based Feedback Control to Accelerate Human Learning ⋆

Madeleine S. Yuh * **Sooyung Byeon** * **Inseok Hwang** *
Neera Jain *

* *Purdue University, West Lafayette, IN 47909 USA (e-mail: myuh, sbyeon, ihwang, neerajain@purdue.edu)*

Abstract: Autonomous systems are increasingly being used for the purpose of training humans to attain new skills or perform new tasks. In these contexts, autonomous systems should be responsive to, and guide, human behavior such that skill or task performance is maximized. These systems generally rely on human performance to determine if assistance is needed. However, it is recognized that these systems should also respond to human cognitive factors, such as self–confidence, that are relevant for human learning. We propose and experimentally validate a heuristic control strategy, based on both a user's performance and self-reported self-confidence as they, that determines whether or not they receive automated assistance in learning how to land a quadrotor in a simulated environment. Through a human subject study involving a benchmark strategy that is solely performance-based, we show that the proposed strategy not only successfully calibrates the self-confidence of the participants, but also leads to statistically significant improvements in participants' task performance and consistency after 20 trials relative to outcomes for participants who experience the benchmark strategy.

Keywords: Co-Learning and self-learning, Human centered automation, Human self–confidence, Shared control

NOMENCLATURE

γ	Change in performance threshold
ϕ	Performance threshold
σ	Self–confidence threshold
θ_k	Landing attitude
d_{max}	Maximum distance possible between quadrotor and landing pad
k	Trial number
$L_{k,n}$	Landing type at trial k, $n \in \{1, 2, 3\}$ for unsuccessful, unsafe, and safe landing respectively
$M_{k,n}$	Control mode at trial k, $n \in \{1, 2\}$ for manual and shared control respectively
$S_{k,\theta}$	Attitude score out of 100 points
$S_{k,p}$	Position score out of 100 points
$S_{k,t}$	Time score out of 100 points
$S_{k,v}$	Velocity score out of 100 points
S_k	Overall score out of 1000 points
SC_k	Self-confidence at trial k
v_k	Landing velocity
x_k	Landing x coordinate
y_k	Landing y coordinate

1. INTRODUCTION

Interactions between humans and automation continue to grow in ubiquity and complexity, in contexts including autonomous vehicles (Knight, 2021; Barfield and Dingus, 2014), military operations (Feickert et al., 2018; Franke, 2014), and product design and manufacturing (Ma et al., 2019). In some contexts, automation is specifically being used for the purpose of training humans in attaining new skills or performing new tasks (Manzey et al., 2011; Kanumuri et al., 2008). Typically, performance-based feedback is used to adapt the automation to the human (Wright et al., 2018; Kaber and Endsley, 2004). Existing systems, e.g., intelligent tutoring systems (ITS), already rely on human performance feedback to predict decision making behavior (Woolf, 2008; Wright et al., 2018; Kaber and Endsley, 2004). However, it has also been shown that cognitive factors, including self-confidence, play an important role in the design of effective human-automation interaction (HAI) (Peters et al., 2015; Hussein et al., 2020; Lee and See, 2004; Gao and Lee, 2006) and how humans learn (Woolf, 2008; Akbari and Sahibzada, 2020; Arroyo et al., 2009; Tao et al., 2020).

Current ITSs select teaching strategies based on student needs, such as improving self-confidence in addition to responding to performance (Akbari and Sahibzada, 2020; Arroyo et al., 2009). Despite this, however, cognitive-state based feedback is not used to adapt automation to the human in learning contexts *outside of the classroom.* Accomplishing this requires an understanding of how self-confidence evolves in HAI contexts outside the classroom, and in turn, algorithms that adapt the automation accordingly. While the dynamics of self-confidence have been studied (Gao and Lee, 2006; Lee and See, 2004), there has been comparatively little research on how to close-the-loop between human and automation. However, existing work by (Roll et al., 2011) has shown that improving stu-

⋆ This material is based upon work supported by the National Science Foundation under Award No. 1836952.

10.1016/j.ifacol.2023.01.111

dent self–regulation skills by calibrating self–assessment to performance in an in–classroom context improves students' capability to learn independently. In other words, automation assistance that is designed to calibrate the self-confidence of the human in real time has the potential to improve learning outcomes.

Therefore, in this paper we propose and validate a heuristic strategy that considers both self-confidence and user performance to determine the need for autonomous intervention in a learning task. Through a human subject study involving a benchmark strategy that is solely performance-based, we show that the proposed strategy not only successfully calibrates the self-confidence of the participants, but also leads to statistically significant improvements in participants' task performance and consistency after 20 trials relative to outcomes for participants who experience the benchmark strategy.

2. HEURISTIC STRATEGY

In order to use human self-confidence information to improve the way automation adapts to a human user, we propose a heuristic control strategy that considers the human's self-reported self-confidence and performance on a trial basis. We assume that the context is one involving a human learning a new skill through repeated trials, in which they can either execute the skill manually or with assistance from an autonomous aid. We define these as manual mode M_1 and shared control mode M_2, respectively. This heuristic strategy aims to calibrate the user's self-confidence to their performance. If self–confidence is high relative to the threshold σ ($SC_k > \sigma$), performance is consistently low relative to the threshold γ ($\Delta S_k \leq \gamma$), and performance is low relative to threshold ϕ ($S_k \leq \phi$), the participant is likely to be over–confident. In this case, they are assigned manual mode (M_1) for the subsequent trial, followed by two trials in shared control mode (M_2), as determined by the conditions shown in Table 1. The logic behind this approach is that the participant will perform better in the two trials of shared control mode than in the one trial of manual mode, thereby demonstrating to the participant that they are still in need of assistance. This should, in turn, decrease their self–confidence. If self–confidence is low and performance is consistently high, the participant is likely under–confident. In this case, two trials of manual mode (M_1) are given, with the logic that the participant will recognize that they are performing well without assistance. This aims to increase the participant's self–confidence. On the other hand, the participant's performance is calibrated if their self-confidence is low when performance decreases or is consistently low, and if their self–confidence is high when when performance improves or is consistently high. In this case, the strategy allocates shared (M_2) and manual (M_1) control to the next trials for low and high self–confidence respectively. It should be noted that aiming to maximize self–confidence is not ideal for task performance (Lee and Moray, 1994). This may lead to mis–calibration of self–confidence and consequently the misuse of the automation (Lee and See, 2004).

As a baseline, we consider a performance-only based heuristic for determining whether the autonomous aid should intervene. If the participant achieves a low score, they are given assistance for the next two trials. If they achieve a high score, manual mode is assigned to the next two trials.

3. HUMAN SUBJECT STUDY

3.1 *Participants*

Forty participants completed the human subject study (17 male and 22 female). Participants were randomly placed in two groups, resulting in 20 participants per group. Participant ages ranged between 18–57 years (mean = 24 years). Each participant was compensated at a rate of \$20/hr. Institutional Review Board at Purdue University approved the study.

3.2 *Experimental Design*

We designed a between subject human study in which participants practice landing a quadrotor in a simulated training module. In the experiment, the participant's goal is to learn to manually land the quadrotor within 20 trials. Participants control the quadrotor using a Thrustmaster T.Flight Hotas controller, as shown in Figure 1, in which the throttle controls the thrust force of the quadrotor and the joystick controls the quadrotor attitude (tilt angle).

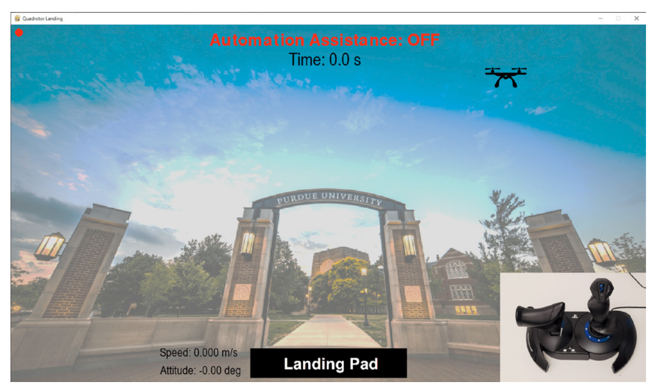

Fig. 1. The quadrotor landing experimental platform.

The training module is equipped with a shared control mode in which they are assisted by a static control law that augments the human's input to successfully land the quadrotor. In each trial, participants practice landing the quadrotor in either manual mode M_1 or shared control mode M_2. The quadrotor input in shared control mode is a convex combination of the human input u_h and the automation input u_a such that the input to the

Table 1. Heuristic strategy using performance metrics and self-confidence cognitive feedback

	Performance Change			
	$\lvert\Delta S_k\rvert > \gamma$ $\Delta S_k < 0$	$\lvert\Delta S_k\rvert \leq \gamma$ $S_k \leq \phi$	$\lvert\Delta S_k\rvert \leq \gamma$ $S_k > \phi$	$\lvert\Delta S_k\rvert > \gamma$ $\Delta S_k > 0$
$SC_k \leq \sigma$	2 M_2	2 M_2	2 M_1	2 M_1
$SC_k > \sigma$	2 M_2	1 $M_1 \rightarrow$ 2 M_2	M_1	2 M_1

Table 2. Heuristic strategy using performance metrics.

$S_k \leq \phi$	$S_k > \phi$
2 M_2	2 M_1

quadrotor is given by $u(n) = 0.9u_h(n) + 0.1u_a(n)$ at each time step n. The automation input of the shared control input was developed by Byeon et al. (2021) using an expert control strategy represented by a linear state feedback controller. The weights of the convex combination were adjusted based on pilot studies such that the task's level of difficulty was deemed appropriate. There are two groups of participants. Group 1 uses the heuristic strategy that is responsive to the user's self–confidence and performance, as depicted in Table 1; we will refer to this as the "confidence strategy". Group 2 uses the benchmark heuristic strategy that responds only to the user's performance, as seen in Table 2; we will refer to this as the "benchmark".

3.3 Score Quantification

Given that both control strategies utilize performance-based feedback, a performance metric, or score, is defined as a function of the context-specific metrics required to improve task performance (Sarangi and Shah, 2015; Scheider et al., 2015; Faghihi et al., 2014). As a result, the numerical score for each trial k is calculated using six performance metrics: time expended per trial t_k, root mean square error RMS between the participant's and expert's trajectory, final position coordinates x_k and y_k, landing velocity v_k, and landing attitude θ_k. Using these performance metrics, sub–scores out of 100 for time, landing position, landing velocity and landing attitude are respectively given by

$$S_{k,t} = \begin{cases} 100\left(1 - \dfrac{1.04(t_k - 5.0)}{1 + e^{\frac{-((t_k-5.0)-45.0)}{15.0}}}\right) & \text{, land} \\ 50\left(1 - \dfrac{RMS - 1.25}{5.0}\right) & \text{, crash} \end{cases} \tag{1}$$

$$S_{k,p} = \begin{cases} 100 & \text{, land} \\ 100\left(1 - \dfrac{x_k^2 + y_k^2}{d_{max}}\right) & \text{, crash} \end{cases} \tag{2}$$

$$S_{k,v} = 100\left(1 - \frac{1}{1 + e^{\frac{-(v_k-8.5)}{2.0}}}\right) \tag{3}$$

$$S_{k,\theta} = 126.4\left(1 - \frac{1}{1 + e^{\frac{-|\theta_k|-20.0)}{15.0}}}\right) . \tag{4}$$

The final numerical score is the sum of the four sub–score equations and is scaled such that a perfect score is 1000 points. The score equations were adjusted using pilot data to ensure that the final score is representative of participants' performance. For example, Sigmoid functions are used to shape the individual scoring functions such that participants who are unsuccessful at landing the quadrotor but are close to achieving safe landing conditions are not penalized as severely as those who lose control of the quadrotor. In the final pilot study, the 50% quantile values from all trials and participants for score, change in score, and numerical self–confidence data were calculated and used as performance and self-confidence thresholds for both heuristic strategies.

3.4 Procedure

After instruction, participants are given two 60–second tutorials to familiarize themselves with the simulator environment. This is followed by 20 trials of the quadrotor game. After every trial, participants are provided with their numerical score, the amount of time they expended in landing the quadrotor, and whether they *unsuccessfully*, *unsafely*, or *safely* landed the quadrotor for all previous trials. Participants achieve a safe landing by landing the quadrotor on the landing pad at a speed less than or equal to 5 m/s and an attitude between −10 and +10 degrees. An unsafe landing occurs when the quadrotor lands on the landing pad but outside the safe landing conditions. An unsuccessful landing is given when the quadrotor crashes outside the landing pad. Additionally, the participant is asked to rate their self–confidence in their ability to land the quadrotor on a numerical scale of 0-100. The survey also asks the participant to rate their trust in the automation, but this data is not utilized within the scope of this paper. Definitions for trust and self–confidence are provided to all participants as follows. The definition of self–confidence is: "*The confidence in oneself and one's powers and abilities.*" The definition of trust is: "*The assured reliance on the character, ability, strength, or truth of someone or something.*" Note that the first two, and last five, trials are completed in manual mode so that each participant's change in manual performance can be quantified. In trials 3–15, the participant's control mode is determined by one of the two heuristic strategies.

4. RESULTS AND DISCUSSION

The resulting data can be analyzed and quantified in several ways to validate the proposed cognitive state-based feedback strategy. We first investigate and compare the calibration of self–confidence between groups 1 and 2. Then, to compare performance between the two groups, we quantify each participant's skill level after 15 trials based upon their score in the last 5 trials of manual performance. Finally, we investigate contributing factors to self–confidence in both groups. The significance of regression variables are evaluated using the t–test p–values and the estimated coefficients.

4.1 Self-Confidence Calibration

We first validate that the confidence strategy achieved the intended goal of calibrating participants' self-confidence. To do this, the median number of trials in which confidence was calibrated is compared between the two groups of participants. Self–confidence is considered calibrated when it increases with improved performance and decreases with deteriorating or consistently low performance. Confidence calibration is assessed using data from trials 2–20 because the change in performance between trials is used to determine if self–confidence is calibrated. The results are shown in Figure 2. Participants in group 1 achieve a higher number of trials with calibrated self–confidence than those in group 2. It should be noted that error–bars are included in Figure 2 but are sufficiently small that they are almost not visible.

The confidence strategy is further validated through independent t–tests. Using numerical score S_k as the response variable and group as a categorical variable, the 20 participants in group 1 who used the confidence strategy ($\mu = 731.54$, $\sigma = 251.94$) compared to those in group 2 who used the benchmark ($\mu = 634.09$, $\sigma = 271.63$)

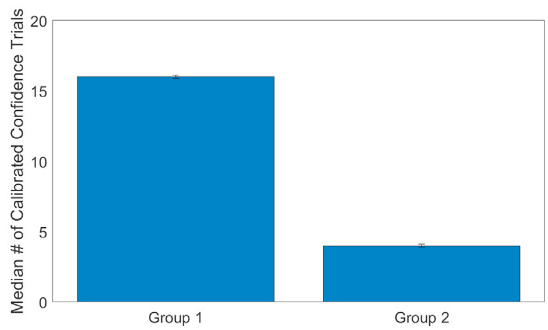

Fig. 2. Confidence calibration comparison between groups. Group 1 participants had more trials with calibrated self–confidence than participants in group 2.

achieved significantly higher scores ($t(753.75) = 5.7063$, $p = 1.659 \times 10^{-8}$). Using numerical self–confidence as the response variable and group as a categorical variable, the 20 participants in group 1 ($\mu = 54.57$, $\sigma = 28.92$) compared to those in group 2 ($\mu = 43.75$, $\sigma = 31.61$) had significantly higher self–confidence ($t(752.11) = 4.9213$, $p = 1.057 \times 10^{-6}$). A power analysis was completed for both independent t–tests to ensure that a type 2 error was not obtained.

From these results, it can inferred that the confidence strategy, which is designed to calibrate self–confidence, is working as intended. Furthermore, these results provide evidence that a heuristic strategy based on both the user's cognitive state and performance can lead to better task performance after a fixed number of trials as compared to a strategy that only considers the user's performance.

4.2 Analysis of Participant Performance

Next, performance is compared between the two groups to determine if the cognitive state-based strategy accelerates learning. Recall that the last five trials are completed in manual mode so that each participant's skill level can be quantified after the training is complete. Note that only data from participants who were able to land unsafely or safely at least three times within their last five trials were used. For groups 1 and 2, 15 and 12 participants satisfied this criterion, respectively.

The absolute performance scores from the last five trials are used to compare performance between groups. Similarly, the variance of scores from the last five trials is used to compare the consistency in performance between groups. The data is presented in the violin plots shown in Figure 3 and Figure 4, respectively.

The violin plots visualize the probability density of the numerical data for each variable (Hintze and Nelson, 1998). For example, in Figure 3, the wider regions of the violin plots identify the scores that participants from each group are likely to achieve. Within each violin plot is a box plot to show the median, interquartile range, maximum, minimum, and outliers of the numerical data. It can be observed from Figure 3 that participants in group 1 are more likely to achieve a higher score than those in group 2. Furthermore, participants in group 1 who safely landed the quadrotor on the landing pad were more consistent in their performance than those in group 2, as shown in Figure 4.

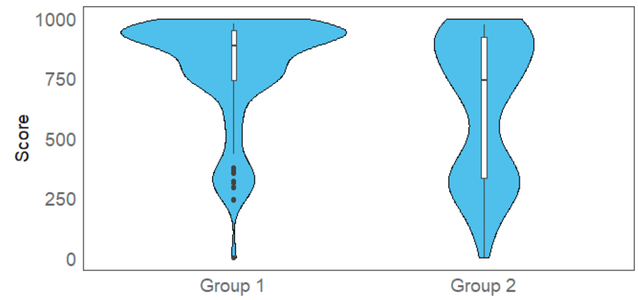

Fig. 3. The comparison of absolute performance out of 1000 points between groups. Participants in group 1 are more likely to achieve higher scores than those in group 2.

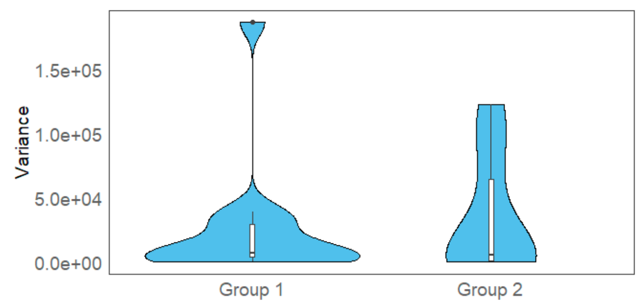

Fig. 4. Performance variance comparison between groups. Group 1 participants are more likely to have lower variance in their final score in comparison to participants in group 2.

These findings show that the confidence strategy leads to increased performance compared to the benchmark.

We can further analyze the performance of participants based on their completion of the quadrotor landing tasks. The participant's goal is to learn how to land the quadrotor safely. Consequently, the efficacy of the two heuristic strategies can be compared by analyzing how many times participants were successful at landing the quadrotor. The number of unsuccessful, unsafe, and safe landing types in each group across all 20 trials is shown in Figure 5. It can be observed that participants in group 1 safely land the quadrotor more frequently in the in the last five trials than those in group 2. Of particular interest is the fact that the point at which safe landings outnumber unsuccessful landings occurs sooner for participants in group 1 than in group 2. This suggests that participants in group 1 are learning how to safely land the quadrotor *faster* than those in group 2, and similarly could be transitioning from novice to expert faster as well.

4.3 Contributing factors to self–confidence:

Finally, to further evaluate the efficacy of the proposed cognitive state-based strategy, multivariate linear regression is used to determine which factors contribute to self-confidence. Participant data from trial 2–20 is utilized for regression analysis as self–confidence of the previous trial SC_{k-1} is used as a regression variable. Two models are created, one for each group of participants. The coefficient estimates and p–values are provided in Figure 6 and Table 3, respectively.

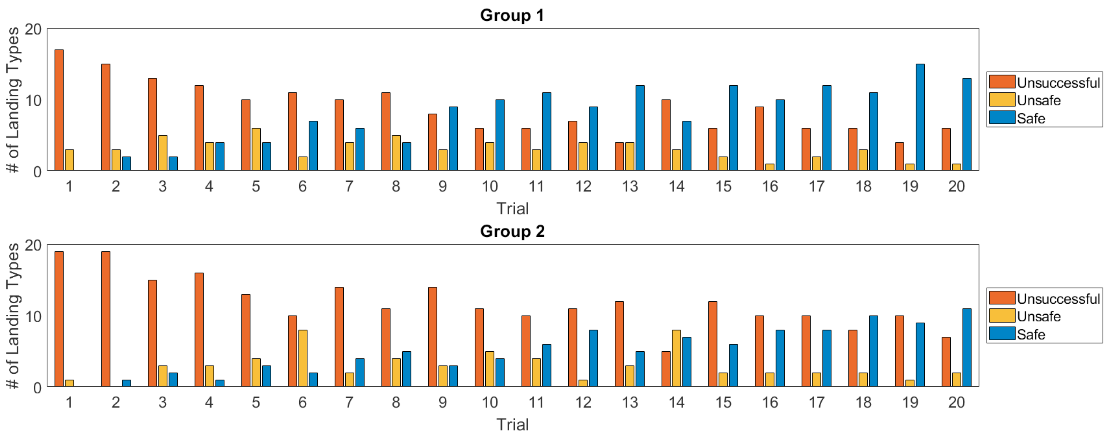

Fig. 5. Bar plot showing number of unsuccessful, unsafe, and safe landings in group 1 and 2 over 20 trials.

Table 3. p-values and significance for self–confidence regression model. Recall that the analysis uses participant data from trials $k \in [2, 20]$. Note: $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Regressor	p–values and Significance Group 1		Group 2	
Intercept	0.750		0.387	
k	0.349		0.900	
SC_{k-1}	$< 2e-16$	***	$< 2e-16$	***
$M_{k,2}$	0.008	**	0.465	
RMS	0.075		0.045	*
$L_{k,1}$	$1.2103 - 05$	***	0.001	**
$L_{k,2}$	$2.690e-04$	***	$7.750e-05$	***
S_k	0.011	*	0.144	
x_k	0.205		0.313	
y_k	0.512		$6.140e-05$	***
v_k	0.059		0.331	
θ_k	0.402		0.337	
t_k	0.036	*	0.383	
Multiple R^2	0.8479		0.8497	
Adjusted R^2	0.8429		0.8448	

From Table 3, for both groups, previous self–confidence SC_{k-1}, unsuccessful landing type $L_{k,1}$, and unsafe landing type $L_{k,2}$ are all significant. From Figure 6, it follows that for both groups, unsuccessful and unsafe landings have the largest predicted effect on self–confidence. For participants in group 1, Score S_k, landing speed v_k, and time per trial t_k are significant. For those in group 2, the root mean square error RMS and landing position y_k are significant. One interpretation of these results is that group 1 participants develop sufficient skill in controlling the quadrotor such that they are able to focus on advanced metrics such as landing speed and how quickly they land, whereas group 2 participants require the majority of the 20 trials to simply develop the skill of flying the quadrotor to the landing pad.

Additionally, shared control mode $M_{k,2}$ is more significant in group 1 than in group 2, and has a positive coefficient. Considering the analysis from Section 4.1, the positive correlation may be due to participants using shared control mode to learn how to better execute the task. This may further indicate that participants in group 1 are learning more from shared control than those in group 2. It should be noted that all variables identified as significant in each regression model have 95% confidence interval coefficient ranges that do not include zero, meaning that the significant regression variables are likely to have some effect on self–confidence.

5. CONCLUSION

We designed and validated a heuristic strategy aimed toward calibrating human self–confidence during human interactions with autonomous in learning contexts. The heuristic strategy utilized both the user's self-reported self-confidence and task performance to determine whether automated assistance would be given to the user at subsequent trials, with an underlying goal of mitigating the adverse effects of over- or under-confidence. We conducted a between subjects study with 40 participants to validate the proposed strategy and compare it against a baseline approach that considers only the user's performance. We showed that participants who received the self–confidence-based strategy were more likely to successfully land the quadrotor manually in the last 5 trials of the experiment, as well as demonstrate consistency in their performance (relative to participants who experienced the baseline strategy). We also showed that participants who used the confidence strategy improved their skill to the point of being able to focus on more advanced landing features, while those who used the baseline strategy may have only focused on flying the quadrotor to the landing pad. This suggests that automation designed for learning contexts should respond to human cognitive state feedback with the intention of calibrating cognitive states, such as self–confidence, to task performance. Future work will be aimed at understanding points of transition from novice to expert, and the design of cognitive state-based strategies that can accelerate such transitions. Furthermore, the proposed heuristic strategy should be implemented for other learning contexts to evaluate its generalizability.

REFERENCES

Akbari, O. and Sahibzada, J. (2020). Students' Self-Confidence and Its Impacts on Their Learning Process. *American International Journal of Social Science Research*, 5(1), 1–15. doi: 10.46281/aijssr.v5i1.462.

Arroyo, I., Cooper, D.G., Burleson, W., Woolf, B.P., Muldner, K., and Christopherson, R. (2009). Emotion Sensors Go To School. In *Proceedings of the 2009 Conference on Artificial Intelligence in Education: Building Learning Systems That Care: From Knowledge Representation to Affective Modelling*, 17–24. IOS Press, NLD.

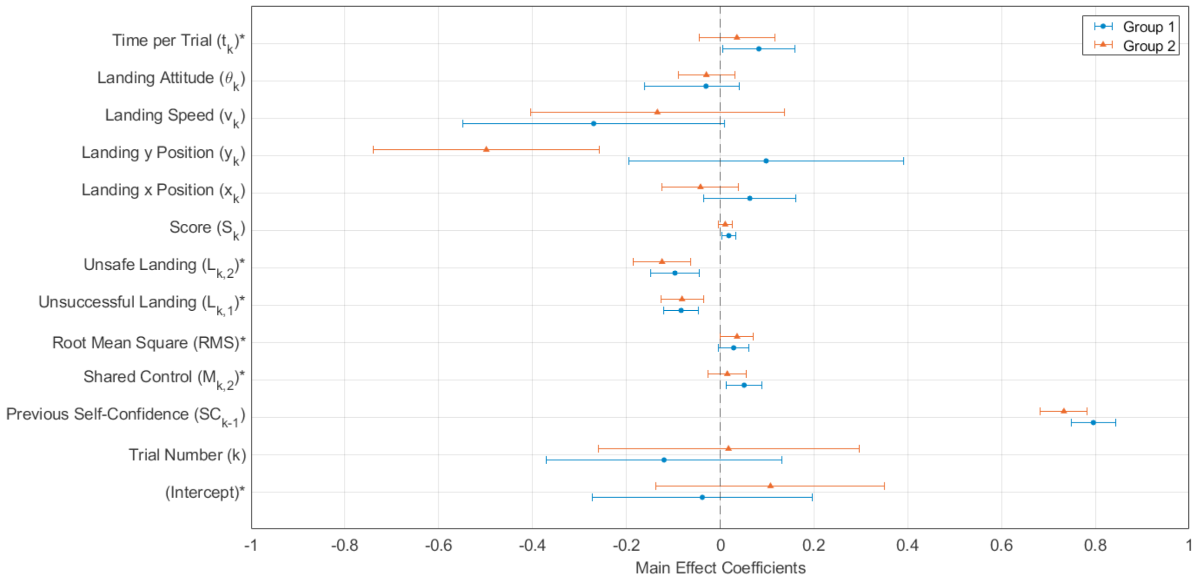

Fig. 6. Plot of variable coefficients in the linear regression model. The estimated coefficients of the regression variables are the circles and triangles for groups 1 and 2, respectively. The ranges are the 95% confidence intervals. Note: * denotes coefficients that have been scaled by a factor of 10^{-2} for improved figure interpretability.

Barfield, W. and Dingus, T.A. (2014). *Human Factors in Intelligent Transportation Systems.* Taylor & Francis, New York, first edition.

Byeon, S., Sun, D., and Hwang, I. (2021). Skill-level-based Hybrid Shared Control for Human-Automation Systems. In *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 1507–1512.

Faghihi, U., Brautigam, A., Jorgenson, K., Martin, D., Brown, A., Measures, E., and Maldonado-Bouchard, S. (2014). How Gamification Applies for Educational Purpose Specially with College Algebra. *Procedia Computer Science*, 41, 182–187.

Feickert, A., Kapp, L., Elsea, J.K., and Harris, L.A. (2018). U.S. Ground Forces Robotics and Autonomous Systems (RAS) and Artificial Intelligence (AI): Considerations for Congress. Congressional Research Service report, Congressional Research Service.

Franke, U.E. (2014). Drones, Drone Strikes, and US Policy: The Politics of Unmanned Aerial Vehicles. *Parameters*, 44(1), 121–130.

Gao, J. and Lee, J.D. (2006). Extending the decision field theory to model operators' reliance on automation in supervisory control situations. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 36(5), 943–959.

Hintze, J.L. and Nelson, R.D. (1998). Violin Plots: A Box Plot-Density Trace Synergism. *The American Statistician*, 52(2), 181.

Hussein, A., Elsawah, S., and Abbass, H. (2020). Towards Trust-Aware Human-Automation Interaction: An Overview of the Potential of Computational Trust Models. In *Proceedings of the 53rd Hawaii International Conference on System Sciences.* Manoa, Hawaii.

Kaber, D.B. and Endsley, M.R. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science*, 5(2), 113–153.

Kanumuri, P., Ganai, S., Wohaibi, E.M., Bush, R.W., Grow, D.R., and Seymour, N.E. (2008). Virtual Reality and Computer-Enhanced Training Devices Equally Improve Laparoscopic Surgical Skill in Novices. *JSLS : Journal of the Society of Laparoendoscopic Surgeons*, 12(3), 219–226.

Knight, W. (2021). Why Tesla Is Designing Chips to Train Its Self-Driving Tech. https://www.wired.com/story/why-tesla-designing-chips-train-self-driving-tech/.

Lee, J.D. and Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies*, 40(1), 153–184.

Lee, J.D. and See, K.A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46(1), 50–80.

Ma, X., Tao, F., Zhang, M., Wang, T., and Zuo, Y. (2019). Digital twin enhanced human-machine interaction in product lifecycle. *Procedia CIRP*, 83, 789–793.

Manzey, D., Luz, M., Mueller, S., Dietz, A., Meixensberger, J., and Strauss, G. (2011). Automation in Surgery: The Impact of Navigated-Control Assistance on Performance, Workload, Situation Awareness, and Acquisition of Surgical Skills. *Human Factors*, 53(6), 584–599.

Peters, J.R., Srivastava, V., Taylor, G.S., Surana, A., Eckstein, M.P., and Bullo, F. (2015). Human Supervisory Control of Robotic Teams: Integrating Cognitive Modeling with Engineering Design. *IEEE Control Systems Magazine*, 35(6), 57–80.

Roll, I., Aleven, V., McLaren, B.M., and Koedinger, K.R. (2011). Metacognitive practice makes perfect: Improving students' self-assessment skills with an intelligent tutoring system. In *International conference on artificial intelligence in education*, 288–295. Springer.

Sarangi, S. and Shah, S. (2015). Individuals, teams and organizations score with gamification: Tool can help to motivate employees and boost performance. *Human Resource Management International Digest*, 23(4), 24–27.

Scheider, S., Martin, R., Kiefer, P., Sailer, C., and Weiser, P. (2015). Score design for meaningful gamification. In *Conference: CHI'15 - Gamifying Research: Strategies, Opportunities, Challenges and Ethics.* Seoul, Korea.

Tao, Y., Coltey, E., Wang, T., Alonso, M., Shyu, M.L., Chen, S.C., Alhaffar, H., Elias, A., Bogosian, B., and Vassigh, S. (2020). Confidence Estimation Using Machine Learning in Immersive Learning Environments. In *2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, 247–252.

Woolf, B.P. (2008). *Building Intelligent Interactive Tutors: Student-centered Strategies for Revolutionizing e-Learning.* Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Wright, J.L., Chen, J.Y.C., and Barnes, M.J. (2018). Human–automation interaction for multiple robot control: The effect of varying automation assistance and individual differences on operator performance. *Ergonomics*, 61(8), 1033–1045.