

# Exploring Interdependencies between Self-Confidence, Workload, and Learning Stage for Intelligent Tutoring Systems

Henry J. Lee

*Department of Computer Science  
Purdue University  
West Lafayette, IN, USA  
lee4602@purdue.edu*

Madeleine S. Yuh

*Department of Mechanical Engineering  
Purdue University  
West Lafayette, IN, USA  
myuh@purdue.edu*

Neera Jain

*Department of Mechanical Engineering  
Purdue University  
West Lafayette, IN, USA  
neerajain@purdue.edu*

**Abstract**—Intelligent tutoring systems (ITS) have been utilized to assist humans in learning new skills in psychomotor learning. Self-confidence, workload, and learning stage are three important factors that contribute to psychomotor learning. Self-confidence and workload have been shown to influence learning outcomes, and psychomotor task performance can be measured in terms of the learning stage. To design ITS algorithms, models will need to capture interactions between relevant cognitive stages and learning outcomes. Previous works utilized reinforcement learning methods by training self-confidence state, workload state, and learning stage state Markov decision process (MDP) models to improve ITS's assistance capability, and the models were trained with the assumption that the states were independent from one another. However, there may be cases where the states could be interdependent (e.g., the relationship between self-confidence and learning progression), or a different model can be explored entirely. In this work, the MDP framework is modified to capture and investigate the interdependencies between different cognitive states, including self-confidence, workload, and learning stage. Our findings suggest that coupling self-confidence with workload or learning stage may enhance the model's predictive capabilities; however, the increased number of model parameters in these coupling states could lead to overfitting. Additionally, we investigate how the models could be improved through k-means clustering, resulting in multiple models for each state to capture a range of human cognitive state behavior.

**Index Terms**—Intelligent Tutoring Systems, Cognitive State Modeling, Markov Decision Process, Reinforcement Learning

## I. INTRODUCTION

### A. Motivation and Problem Statement

Adaptive automation, defined as automation that reacts to human behavior through control laws or assistance [20], is a crucial tool for helping humans train when learning and acquiring new skills in dynamic systems. Automation has been used in several fields, including manufacturing [16], product design [8], military applications [3], and healthcare [4], [6]. In particular, there is a growing interest in developing adaptive automation for tutoring or training systems, also

known as intelligent tutoring systems (ITS). When paired with cognitive feedback, adaptive automation has the potential to oversee human training but requires precise assessments of human progression through different learning stages [5], [17]. Feedback is important to improve performance [15] and this can be achieved through formative feedback, which informs users of their progress during the learning process, thereby enhancing their performance [14].

### B. Background

Intelligent tutoring systems (ITS) are defined as computer environments that teach students through adaptive algorithms that can accommodate the needs of a human learner in terms of knowledge, cognition, motivation, and sometimes emotions [10]. An ITS has the adaptivity that can consider an operator's previous performance, a model of the subject, and other psychological characteristics [10]. In existing literature, it is recognized that human cognitive states are important to understand and account for in human automation interaction systems [11]. Prior works investigated the role of human self-confidence and workload in psychomotor task training, specifically landing a quadrotor in a 2D quadrotor simulator module in intelligent tutoring systems. ITS development for psychomotor tasks often faces three challenges: (1) creating a knowledge space for the task, (2) maintaining learner motivation, and (3) personalization of the tutoring agent to learner characteristics [19], [18]. To address all three challenges, a landing quadrotor simulator was implemented with automation that responds to learner self-confidence, workload, and learning stage.

Self-confidence is an important component when it comes to human-machine interactions. Self-confidence is defined as "one's judgment about their capability to accomplish a task or goal" [2], [19]. Over-confidence and under-confidence can undermine learning, which requires a need to calibrate self-confidence and motivation to match their skill level [20]. An experiment involving a landing quadrotor simulator examined self-confidence by examining the trajectories of the landings and putting those trajectories in four different skill levels [20]

This material is based upon work supported by the National Science Foundation under Awards No. 1836952. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

and implemented automation to assist the user in landing the quadrotor in four possible control objectives: maximizing performance, maximizing self-confidence, calibrating self-confidence to performance, and maximizing both performance and self-confidence [19]. A review of self-confidence suggested that when designing an automation system, the operator's responsibilities should be defined by the operator's capabilities rather than how the automation is implemented [11].

Workload is another component when it comes to human-machine interactions. Workload is defined as "the extent to which a task demands on the operator's cognitive resources, with a variety of models further reducing the construct into various subcomponents" [12]. An important task of ITS is to reduce the amount of mental energy a task may require from a human, since mental workload is closely related to operator stress [12]. Stress is interlinked with workload and the automation needs to account how well a learner is doing because too much stress can overwhelm a learner when they learn in a challenging environment while too little stress leads to boredom and vigilance decrease since the learner does not need to use too much mental energy to do the task [12]. This is a fundamental reason for introducing automation into systems since the aim is to reduce the chance of human error by adjusting the mental workload on a learner when doing a task [11].

Cognitive state feedback should be paired with automation to encourage appropriate self-confidence and workload. However, in addition to that, good understanding of human performance is required to measure learning. It is important that the automation adjusts itself to the appropriate level of the learner since it can prevent the human from under-using or overusing the automation. [1] proposed a model that characterizes the acquisition of new skills into 5 stages: novice, advanced beginner, competent, proficient, and expert. Each stage can describe the learning progress of a learner, where a novice learner decomposes and understands a task, an advanced beginner learner develops strategies, identifies sub-tasks, and recognizes instability cues, a competent learner refines strategy, prioritizes sub-tasks, increases consistency, and has high attention load, and a proficient learner has a fixed strategy, improves with practice, and decreased their attention load while doing the task [20]. In this work, the proficient and expert stages is merged together because a learner is more likely to reach the proficient learning stage before reaching the expert learning stage in the landing quadrotor simulator [20].

### C. Gaps in Literature

Prior work in [20] addressed challenge 1 by utilizing a learning stage system. Each trial in the landing quadrotor system tracked a projector, compared it to a master baseline, and categorized as one of novice, advanced beginner, competent, or proficient. [19] addressed challenge 2 by utilizing a Markov Decision Process (MDP) model to calibrate self-confidence with performance. The MDP model is used to calculate performance and self-confidence scores, which are

then used to determine if the automation should be turned on or not. [18] addressed the third challenge by giving structured feedback from GPT-4o. The tone of the feedback can be positive, neutral, or negative, depending on a person's self-confidence and performance levels. This work extended the modeling work in [19] by utilizing learning stage, self-confidence, and workload state MDP models to then train an optimal policy that determined when to assist users as well as tone of feedback.

However, in [18], it was assumed that self-confidence, workload, and learning stage were independent from one another. This may not be true, and the previous model frameworks may not capture the interactions between the three states. Additionally, there is an opportunity to investigate whether or not clustered behavior may exist, requiring multiple models to represent self-confidence, workload, and learning stage behavior.

### D. Contributions

This paper focuses on two specific problems: (1) exploring alternative model frameworks that relax the independent assumption between self-confidence, workload and learning state by training models with coupled states and (2) determining if clustering results in more representative models of self-confidence, workload, and learning stage behavior.

### E. Experiment Framework

This paper utilized a landing quadrotor simulator which was used to measure self-confidence and workload levels in automation [18]. The task of the learner was to learn how to land on the landing pad successfully. A learner can achieve three types of landings: (1) a safe landing which consists of the quadrotor landing on the landing pad with safe landing conditions of a speed less than or equal to 5 m/s and an angle between  $-10^\circ$  and  $+10^\circ$ , (2) an unsafe landing which consists of the quadrotor landing on the landing pad but outside the safe landing conditions, and (3) an unsuccessful landing which occurs when the quadrotor does not land on the landing pad [18]. Figure 1 describes the different areas of interest of the landing quadrotor, which include an automation indicator and time, the quadrotor itself, the live speed and altitude indicator, and the landing pad.

A learner is given 20 attempts to land the quadrotor. After each trial, the learner is asked to self-report their self-confidence and workload scores on a numerical scale from 0 to 100. The learner's trajectory is categorized as learning stage 1 (novice), 2 (advanced beginner), 3 (competent), or 4 (proficient) [20]. Once the self-reported scores are processed, personalized feedback will be given to the learner based on their performance, self-reported scores, and learning stage. The automation of the landing quadrotor simulator will be turned on depending on whether the learner is struggling to land on the landing pad successfully.

### F. Participants

In total, 178 participants (100 male, 78 female) with ages ranging 18-72 years ( $\mu = 33.2$ ,  $\sigma = 12.6$ ) completed the



Fig. 1: The quadrotor landing simulator with four areas of interest: (1) automation indicator and time, (2) quadrotor, (3) live speed and altitude indicator, and (4) landing pad.

user study on Prolific [13]. All participants were compensated \$8/hr. The Institutional Review Board at Purdue University, USA, approved the study.

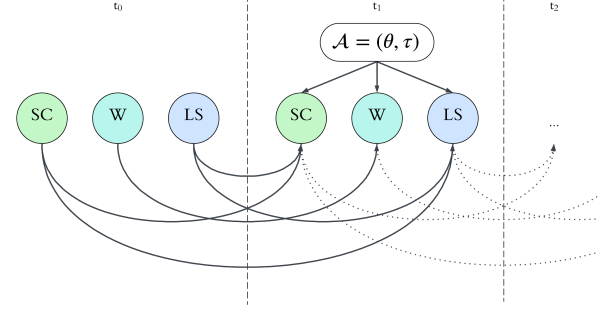
## II. METHODOLOGY

In this section, we describe the methodologies used to (1) explore alternative modeling frameworks through training coupled self-confidence (SC), workload (W), and learning stage (LS) models and (2) determine the ideal number of clusters to divide participant data. All of the code was written in MATLAB and stored in Github Enterprise.

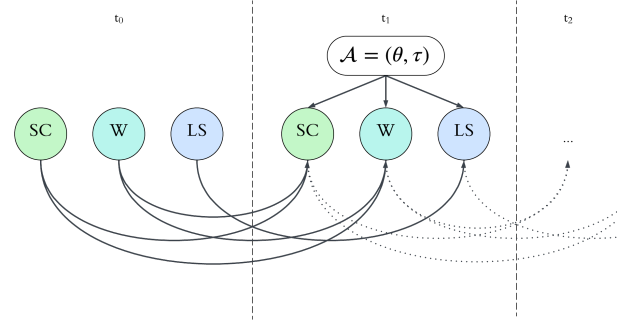
### A. Coupled State Model Exploration

The automation in the landing quadrotor simulator decides when to assist based on an optimal control policy trained using a Markov decision process (MDP) models of self-confidence, workload, and learning stage. An MDP is a framework for describing sequential decision-making problems [19], [18]. Formally, it is defined as a 5-tuple  $(S, \mathcal{A}, \mathcal{T}, \mathcal{R}, \gamma)$ , where  $S$  is a finite set of states,  $\mathcal{A}$  is a finite set of actions,  $\mathcal{T}$  is the probability of transitioning to the next state  $s'$  given the current state  $s$  and action  $a$ ,  $\mathcal{R}(s, a)$  is the reward function that represents the direct rewards associated with the agent's decisions depending on the state and actions, and  $\gamma$  is the discount factor where the value is between 0 and 1 and represents how important future rewards are at the current state [18], [19].

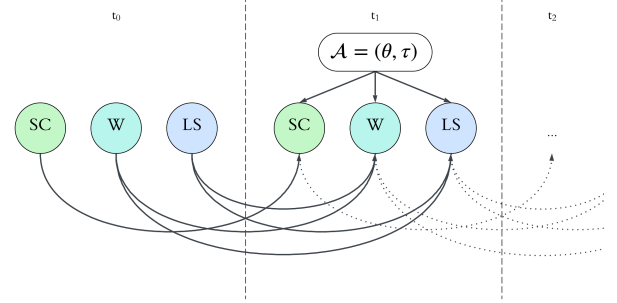
There are three states of interest: self-confidence (SC), workload (W), and learning stage (LS). In prior work, separate MDP models for each state was trained with the assumption that these states were independent from one another. However, to extend the modeling method, we aim to relax the independence assumption and investigate is coupling the states during model training results in more informative models. Therefore, three new models were built and trained as portrayed in Figure 2: (1) a coupled self-confidence and workload model with an independent learning stage model, (2) a coupled self-confidence and learning stage model with an independent



(a) Coupled SC and LS and Independent W Model



(b) Coupled SC and W and Independent LS Model



(c) Coupled W and LS and Independent SC Model

Fig. 2: Coupled models showcasing relationships between SC, W, and LS

workload model, and (3) a coupled workload and learning stage coupled model with an independent self-confidence learning model. Once all of the coupled models were trained, log-likelihood, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and receiver operator characteristic curves were used to examine which coupled model out of the three performed the best, and to examine if the coupled models performed better compared to the independent model. Log-likelihood is a statistical measure used to assess how well a model fits the observed data. Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values are metrics used to compare models [7]. The general idea is that

a model that fits better reduces error and a model that is too complex risks overfitting [7].

After determining which coupled model performed the best, each model (including the independent model) would use ROC curves as seen in 3, which compares the true positive rate to the false positive rate. From this area under the ROC Curve (AUC) values can be used to represent model predictive capabilities. For AUC values, a model performs better if it is closer to 1 than 0.5 since 1 represents a perfect classification while 0.5 represents a random guess.

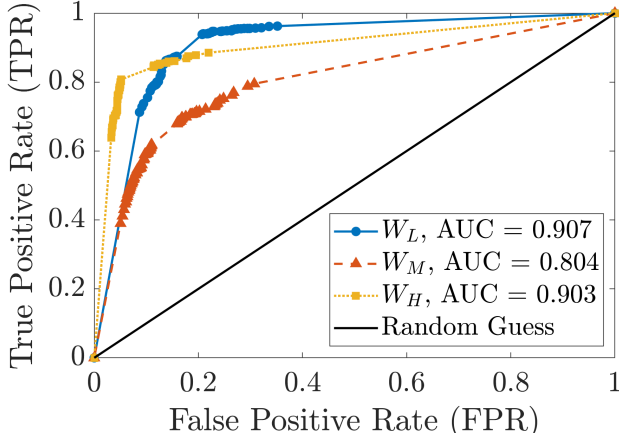


Fig. 3: Resulting ROC curve for the independent workload model

### B. Clustering

For clustering, a  $k$ -means clustering algorithm was used to group participant data into a different number of clusters. BIC values were used as another evaluation metric to determine what is the optimal number of clusters to group the participant data. Algorithm 1 was derived from [9], where it also utilized a clustering algorithm but for partially observable Markov decision process (POMDP) models.

**Algorithm 1** The algorithm used to divide participant into a set of  $k_c$  clusters based on their data. MDP independent models are trained as taken from [9].

- 1: **Result:** A group of  $k_c$  MDP models.
- 2: Create  $k_c$  random clusters using all participants.
- 3: **repeat**
- 4:   Train a set of  $k_c$  MDP models  $\Theta^{k_c}$ .
- 5:   Find the log-likelihood of each participant's data using each model with parameters  $\theta_i^{k_c}$  for  $i \in \{1, \dots, k_c\}$ .
- 6:   Move participants to the cluster whose model has the highest log-likelihood of observing their data given the model parameters.
- 7: **until** No participants are moved between clusters **or** a maximum number of iterations have been completed.

We implemented clustering on the independent SC, W, and LS models. Through this investigation, we aim to determine if

more than one model is necessary to represent human cognitive states and learning stage.

## III. RESULTS AND DISCUSSION

In this section, we explain how (1) the independent model performed better than the coupled models and (2) 2 was the ideal number of clusters to divide self-confidence and workload participant data, and 3 for learning stage participant data.

### A. Coupled State Model Exploration

After training all coupled models, the coupled self-confidence and learning stage, paired with the independent workload learning stage model performed the best out of the three coupled models. This model results in the highest log-likelihood, and lowest AIC and BIC, as seen in Table I. This indicates relaxing the independence assumption between self-confidence and learning stage may be beneficial for capturing the interaction between self-confidence behavior and learning stage.

TABLE I: Comparison of Log-Likelihood, AIC, and BIC Across Different Model Structures

Model Type	Log-Likelihood	AIC	BIC
Coupled SC & W and LS	-8516	17987.75	20940.60
Coupled SC & LS and W	-8142	17239.04	20191.89
Coupled W & LS and SC	-8314	17583.35	20536.21
Coupled SC, W, & LS Model	-11681	31838.32	58018.63
Independent Model	-6840	13907.64	14611.87

However, the independent models (SC + W + LS) performed better than the three coupled models, with it having better overall AUC values as seen in Table II. This is due to the fact that the independent models have a lower number of parameters. The independent model was trained with 38 parameters, and the partially coupled models were trained with 440 parameters. One item to address was the fact that the full coupled model was never used as a comparison. A fully coupled model has 4238 parameters, which is vastly more parameters compared to the independent models. These results indicate that more data may be necessary to fully investigate the relaxation of the independence assumption and avoid overfitting.

### B. Clustering

After the  $k$ -means algorithm was implemented, the data revealed that the optimal number of clusters was 2 for self-confidence, 2 for workload, and 3 for the learning stage, as seen in Figure 4.

TABLE II: Area Under the Curve (AUC) Results Across All Models

(a) AUC Values for Coupled SC and LS and W Independent Model

SC Comparison	AUC	W Comparison	AUC	LS Comparison	AUC
$SC_L$ vs $SC_M$	0.824	$W_L$ vs $W_M$	0.857	$LS_1$ vs $LS_2$	0.553
$SC_L$ vs $SC_H$	0.949	$W_L$ vs $W_H$	0.964	$LS_1$ vs $LS_{3,4}$	0.563
$SC_L$ vs Rest	0.901	$W_L$ vs Rest	0.907	$LS_1$ vs Rest	0.56
$SC_M$ vs $SC_L$	0.797	$W_M$ vs $W_L$	0.83	$LS_2$ vs $LS_1$	0.43
$SC_M$ vs $SC_H$	0.703	$W_M$ vs $W_H$	0.779	$LS_2$ vs $LS_{3,4}$	0.528
$SC_M$ vs Rest	0.742	$W_M$ vs Rest	0.804	$LS_2$ vs Rest	0.486
$SC_H$ vs $SC_L$	0.822	$W_H$ vs $W_L$	0.922	$LS_{3,4}$ vs $LS_1$	0.58
$SC_H$ vs $SC_M$	0.794	$W_H$ vs $W_M$	0.888	$LS_{3,4}$ vs $LS_2$	0.513
$SC_H$ vs Rest	0.809	$W_H$ vs Rest	0.903	$LS_{3,4}$ vs Rest	0.557

(b) AUC Values for Coupled SC W Coupled and LS Independent Model

SC Comparison	AUC	W Comparison	AUC	LS Comparison	AUC
$SC_L$ vs $SC_M$	0.818	$W_L$ vs $W_M$	0.428	$LS_1$ vs $LS_2$	0.671
$SC_L$ vs $SC_H$	0.952	$W_L$ vs $W_H$	0.536	$LS_1$ vs $LS_{3,4}$	0.785
$SC_L$ vs Rest	0.9	$W_L$ vs Rest	0.478	$LS_1$ vs Rest	0.754
$SC_M$ vs $SC_L$	0.795	$W_M$ vs $W_L$	0.585	$LS_2$ vs $LS_1$	0.633
$SC_M$ vs $SC_H$	0.699	$W_M$ vs $W_H$	0.358	$LS_2$ vs $LS_{3,4}$	0.477
$SC_M$ vs Rest	0.738	$W_M$ vs Rest	0.47	$LS_2$ vs Rest	0.544
$SC_H$ vs $SC_L$	0.806	$W_H$ vs $W_L$	0.361	$LS_{3,4}$ vs $LS_1$	0.765
$SC_H$ vs $SC_M$	0.787	$W_H$ vs $W_M$	0.449	$LS_{3,4}$ vs $LS_2$	0.636
$SC_H$ vs Rest	0.797	$W_H$ vs Rest	0.409	$LS_{3,4}$ vs Rest	0.722

(c) AUC Values for Coupled W and LS and SC Independent Model

SC Comparison	AUC	W Comparison	AUC	LS Comparison	AUC
$SC_L$ vs $SC_M$	0.818	$W_L$ vs $W_M$	0.869	$LS_1$ vs $LS_2$	0.587
$SC_L$ vs $SC_H$	0.946	$W_L$ vs $W_H$	0.96	$LS_1$ vs $LS_{3,4}$	0.628
$SC_L$ vs Rest	0.897	$W_L$ vs Rest	0.911	$LS_1$ vs Rest	0.617
$SC_M$ vs $SC_L$	0.802	$W_M$ vs $W_L$	0.82	$LS_2$ vs $LS_1$	0.436
$SC_M$ vs $SC_H$	0.638	$W_M$ vs $W_H$	0.829	$LS_2$ vs $LS_{3,4}$	0.53
$SC_M$ vs Rest	0.706	$W_M$ vs Rest	0.824	$LS_2$ vs Rest	0.49
$SC_H$ vs $SC_L$	0.782	$W_H$ vs $W_L$	0.965	$LS_{3,4}$ vs $LS_1$	0.618
$SC_H$ vs $SC_M$	0.761	$W_H$ vs $W_M$	0.899	$LS_{3,4}$ vs $LS_2$	0.542
$SC_H$ vs Rest	0.772	$W_H$ vs Rest	0.929	$LS_{3,4}$ vs Rest	0.593

(d) AUC Values for Independent Model

SC Comparison	AUC	W Comparison	AUC	LS Comparison	AUC
$SC_L$ vs $SC_M$	0.883	$W_L$ vs $W_M$	0.916	$LS_1$ vs $LS_2$	0.729
$SC_L$ vs $SC_H$	0.967	$W_L$ vs $W_H$	0.968	$LS_1$ vs $LS_{3,4}$	0.830
$SC_L$ vs rest	0.934	$W_L$ vs rest	0.943	$LS_1$ vs rest	0.800
$SC_M$ vs $SC_L$	0.884	$W_M$ vs $W_L$	0.905	$LS_2$ vs $LS_1$	0.694
$SC_M$ vs $SC_H$	0.771	$W_M$ vs $W_H$	0.832	$LS_2$ vs $LS_{3,4}$	0.503
$SC_M$ vs rest	0.805	$W_M$ vs rest	0.867	$LS_2$ vs rest	0.597
$SC_H$ vs $SC_L$	0.828	$W_H$ vs $W_L$	0.906	$LS_{3,4}$ vs $LS_1$	0.797
$SC_H$ vs $SC_M$	0.838	$W_H$ vs $W_M$	0.888	$LS_{3,4}$ vs $LS_2$	0.627
$SC_H$ vs rest	0.834	$W_H$ vs rest	0.897	$LS_{3,4}$ vs rest	0.745

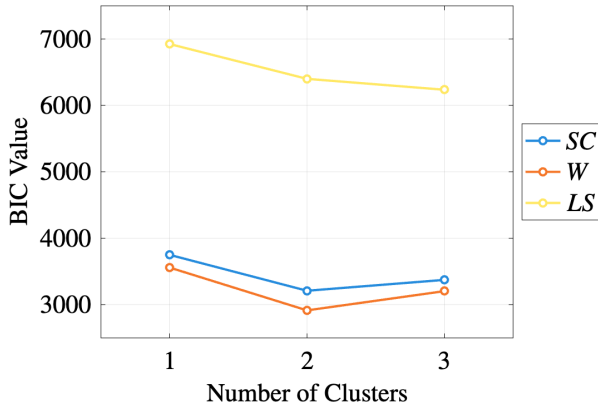


Fig. 4: BIC values collected from clustering independent MDP models

In Figure 5, a heatmap is displayed for a variety of different self-confidence states. The data suggests that for one cluster, learner self-confidence was not easily perturbed (e.g., participants with low self-confidence often stayed with low self-confidence). For the other cluster, self-confidence was more likely to transition between the three levels of low, medium, and high. A similar story can be told with the workload heatmap as shown in Figure 6. These findings are similar to [9], where the optimal number of clusters was 2.

#### IV. CONCLUSIONS

In conclusion, this work explored the interdependencies between self-confidence, workload, and learning stage with intelligent tutoring systems for psychomotor learning tasks. The investigation was split into two problems: (1) examining coupled models and comparing them to the independent model

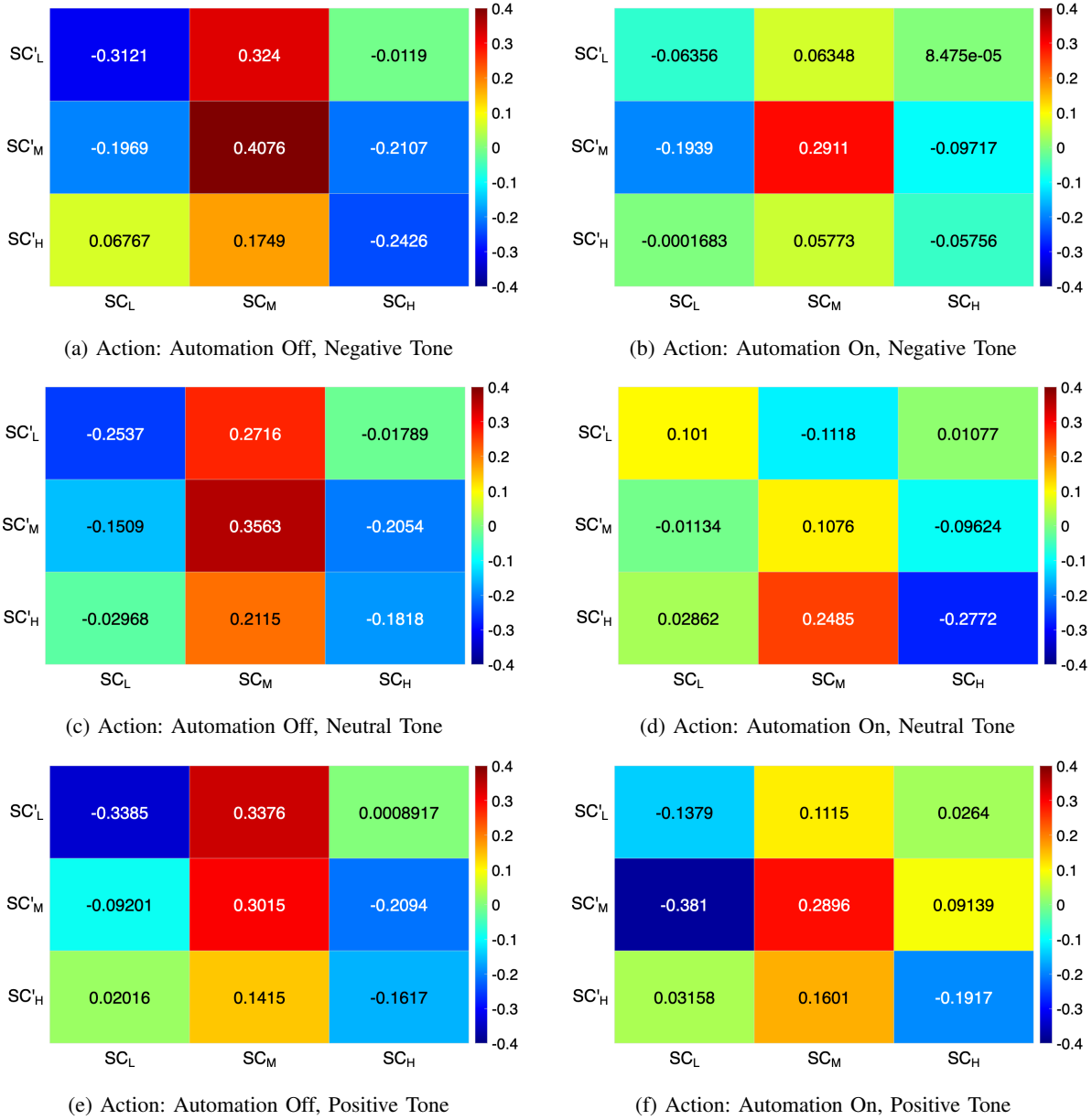


Fig. 5: SC heatmaps depicting differences in transition probabilities between clusters

and (2) determining if clustering participant data into the different groups has a tangible benefit. We trained three coupled models of self-confidence and workload, self-confidence and learning stage, and workload and learning stage to examine how effective they were compared to the independent. We also developed a  $k$ -means clustering algorithm to cluster the participant data from the independent model into  $k$  different clusters.

Our results indicated that the self-confidence and learning stage coupled model was the best-performing coupled model, the independent model performed better than the coupled models, and clustering participant data into 2 or 3 clusters seemed to have a tangible benefit.

Future next steps include policy training for the coupled models and exploring an alternative model to the MDP. Policy training for the coupled model involves utilizing the coupled models to replace the independent model on the landing quadrotor simulator, collecting data for each coupled model on the simulator, and using them as an evaluation metric to determine how well the coupled models performed. An alternative model could replace the MDP in the form of a partially observable Markov decision process (POMDP) model since similar works including [9] utilized POMDPs.

#### ACKNOWLEDGMENT

The authors give thanks to Purdue's Summer Undergraduate Research Fellowship program for sponsoring this program, as



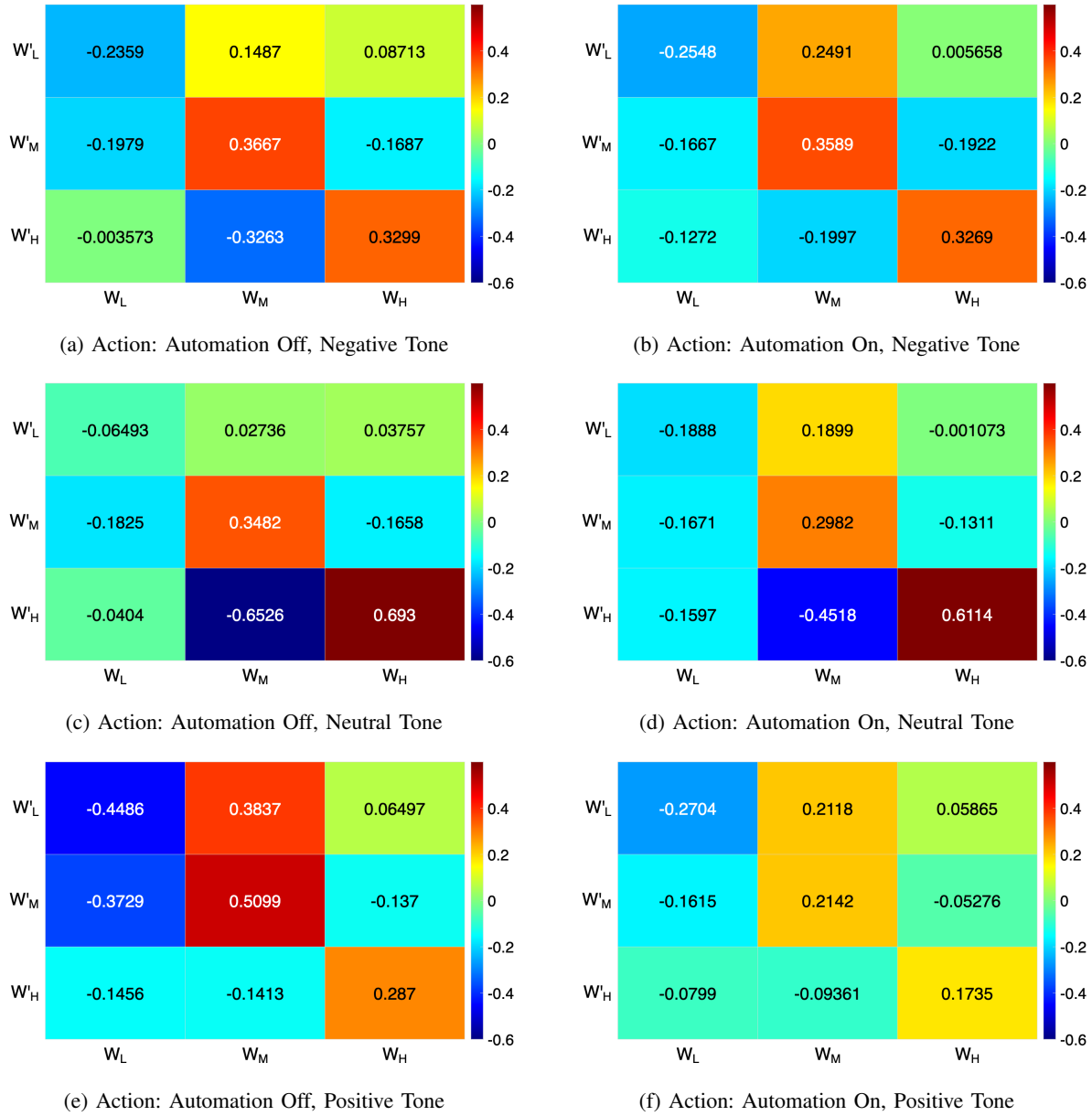


Fig. 6: W heatmaps depicting differences in transition probabilities between clusters

well as acknowledging Varun Aggarwal for being the first author's SURF GA.

## REFERENCES

- [1] H. Dreyfus and S. E. Dreyfus. *Mind over machine*. Simon and Schuster, 1986.
- [2] D. E. Druckman and R. A. Bjork. *Learning, remembering, believing: Enhancing human performance*. National Academy Press, 1994.
- [3] J. K. Elsea, A. Feickert, L. A. Harris, and L. Kapp. U.s. ground forces robotics and autonomous systems (ras) and artificial intelligence (ai): Considerations for congress. CRS Report R45392, Congressional Research Service, Library of Congress, Washington, D.C., Nov. 2018.
- [4] A. J. Ghergich. How automation is transforming healthcare jobs. *Forbes*, Dec. 2021.
- [5] D. B. Kaber and M. R. Endsley. 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, 2004.
- [6] F. Kazzazi. The automation of doctors and machines: A classification for ai in medicine (adam framework). *Future Healthcare Journal*, 8(2):e257–e262, 2021.
- [7] J. Kuha. Aic and bic: Comparisons of assumptions and performance. *Sociological methods & research*, 33(2):188–229, 2004.
- [8] X. Ma, F. Tao, M. Zhang, T. Wang, and Y. Zuo. Digital twin enhanced human-machine interaction in product lifecycle. *Procedia Cirp*, 83:789–793, 2019.
- [9] G. McMahon, K. Akash, T. Reid, and N. Jain. On modeling human trust in automation: Identifying distinct dynamics through clustering of markovian models. *IFAC-PapersOnLine*, 53(5):356–363, 2020.
- [10] B. Morgan, M. Hogan, A. J. Hampton, A. Lippert, and A. C. Graesser. The need for personalized learning and the potential of intelligent tutoring systems. In *Handbook of learning from multiple representations and perspectives*, pages 495–512. Routledge, 2020.

- [11] R. Parasuraman and V. Riley. Humans and automation: Use, misuse, disuse, abuse. *Human factors*, 39(2):230–253, 1997.
- [12] J. R. Peters, V. Srivastava, G. S. Taylor, A. Surana, M. P. Eckstein, and F. Bullo. Human supervisory control of robotic teams: Integrating cognitive modeling with engineering design. *IEEE Control Systems Magazine*, 35(6):57–80, 2015.
- [13] Prolific. Prolific, Mar. 2025.
- [14] V. J. Shute. Focus on formative feedback. *Review of educational research*, 78(1):153–189, 2008.
- [15] A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, and M. Goodrich. Common metrics for human-robot interaction. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 33–40, 2006.
- [16] D. Westrom. The growing use of automation in manufacturing. *MachineMetrics.[Online]*. Available: <https://www.machinemetrics.com/blog/automation-in-manufacturing>, 2020.
- [17] J. L. Wright, J. Y. Chen, and M. J. Barnes. Human–automation interaction for multiple robot control: the effect of varying automation assistance and individual differences on operator performance. *Ergonomics*, 61(8):1033–1045, 2018.
- [18] M. S. Yuh and N. Jain. Designing cognitively-aware psychomotor intelligent tutoring systems via multi-objective and human-centric optimization. In *Proceedings of the IEEE International Conference on Robot and Human Interactive Communication (IEEE RO-MAN)*. IEEE, 2025.
- [19] M. S. Yuh, E. Rabb, A. Thorpe, and N. Jain. Using reward shaping to train cognitive-based control policies for intelligent tutoring systems. In *2024 American Control Conference (ACC)*, pages 3223–3230. IEEE, 2024.
- [20] M. S.-T. Yuh, K. R. Ortiz, K. S. Sommer-Kohrt, M. Oishi, and N. Jain. Classification of human learning stages via kernel distribution embeddings. *IEEE Open Journal of Control Systems*, 3:102–117, 2024.