195

The Effect of Time Delay on Emotion, Arousal, and Satisfaction in Human-Robot Interaction

The Effect of Time Delay on Emotion, Arousal, and Satisfaction in Human-Robot Interaction

Euijung Yang, Michael C. Dorneich
Industrial and Manufacturing Systems Engineering, Iowa State University

This study investigates the influence of time delay on user emotion, arousal, and satisfaction in human-robot interaction (HRI). Time delay is a gap between an input from an operator and the corresponding feedback response from the system, and its negative consequence on performance has been documented in various areas including psychology and HRI. On the contrary, the effects of time delay on user emotion and satisfaction have been difficult to establish due to the fluctuations in the emotional aspect of the physiological state of the operator. In this study, the hypothesis of whether time delay during robot vehicle operation increases user arousal and negative emotions while lowering user satisfaction was tested. Participants were asked to remote-control a robot vehicle to navigate different mazes in a remote location. Time delay was manipulated by introducing lags into system feedback. Subjective and objective measures included emotion tracking through face recognition, and electrodermal activity (EDA). User frustration, anger, and arousal increased while user satisfaction decreased. A better understanding of how time delay influences operator's emotion and how change in emotion is expressed in physiological signals would be of crucial importance to designing an affect-aware robotic systems that have the ability to appropriately respond to user emotional state.

INTRODUCTION

Remote-controlled robot finds many practical applications in areas such as military bomb disposal (Drascic, Milgram, & Grodski, 1989), radioactive environments maintenance (Draper, 1993), surgery (Casals, 1998), mine detection (Nonami, Shimoi, Huang, Komizo, & Uchida, 2000), and subsea manipulation (Ridao, Carreras, Hernandez, & Palomeras, 2007). Despite technological advancements made in in the responsiveness of robotic system, time delay continues to be an issue. Time delays arise during the transmission of information between a robot and an operator (Prewett, Johnson, Saboe, Elliott, & Coovert, 2010; Owen-Hill, Suárez-Ruiz, Ferre, & Aracil, 2014), and is a relatively common problem that arises when interacting across far distances (Corde Lane et al., 2002).

Time delay has been studied in domains ranging from the business settings to internet browsing. Delayed responses from a human being or a computer can increase user stress, aversive behavior, anxiety, impatience, and irritation (Barber & Lucas, 1983; Guynes, 1988; Kuhmann, Boucsein, Schaefer, & Alexander, 1987; Schleifer & Amick, 1989; Szameitat, Rummel, Szameitat, & Sterr, 2009). While time delay is commonly seen as a shortcoming that could be lessened by the use of improved technology, it is also regarded as an unavoidable, inherent component of teleoperation technology and HRI (Adelstein, Lee, & Ellis, 2003; Lum, Rosen, Lendvay, Sinanan, & Hannaford, 2009).

Many studies have focused on the negative consequences of system delay on HRI. However, more empirical studies of system delay on operator's emotions and satisfaction are needed in order build concrete understanding of operator's emotional and arousal reactions when interacting with robots. Such an understanding would be invaluable as it would enable creating a system that could effectively assess and respond to user's emotions to ultimately yield improved human-robot joint performance. Moreover, detecting operator's conditions such as emotional state, body activation, and satisfaction are key to building affect-aware systems to mitigate user's negative states and improve interaction contentment in HRI.

Navigation of a robot vehicle in remotely located mazes was conducted to investigate the emotional, physiological, and usability effects of time delay. Time delay was manipulated by introducing lags into system feedback. The next section discusses related work in time delay issues in HRI and the ways to detect human emotion. The experiment method and results are described. The paper ends with a discussion on the effects of time delay in HRI.

RELATED WORK

Time Delay in Human-Robot Interaction

System lags have been shown to exert negative consequences on diverse aspects of human operator's performance. System lags have been shown to decrease the accuracy of task completion in teleoperation of a robot arm performing a welding task (Owen-Hill et al, 2014). Control performance and accuracy in a positioning task also degraded due to system latency (Corde Lane et al., 2002). System feedback lags has led to reduced perceptual sensitivity of users in virtual environments (Ellis, Mania, Adelstein, & Hill, 2004), increased task error rate (Szameitat, Rummel, Szameitat, & Sterr, 2009), and higher operator workload (Sheik-Nainar, Kaber, & Chow, 2005). Task efficiency improved when lags were eliminated (Luck, McDermott, Allender, & Russell, 2006). For example, observation sensitivity of operator was poorer with system lags. It showed that time delay from interaction processing led to lower human operator's ability to observe environment by using teleoperation of an unmanned ground vehicle on multiple screens (Luck, McDermott, Allender, & Russell, 2006).

Decrease in user satisfaction due to time delays can cause dissatisfied users to seek an alternative system (Hoxmeier & DiCesare 2000). Moreover, system lags cause lower satisfaction of users (Kuhnmann, 1989; Shneiderman, 1992, Hoxmeier & DiCesare 2000; Davis & Hantula, 2001).

Manipulating Emotions

Lags in system feedback have been used to induce frustration in a computer game (Klein, Moon, & Picard, 2002; Hone, 2006). Delays in acting on mouse clicks also induced frustration (Powers, Rauh, Henning, Buck, & West, 2011).

Detecting Human Emotion

Various sensors have been used to assess emotional state, including: electrodermal activity (EDA), event-related brain potentials, electroencephalography, skin conductivity (GSR), palmar sweat, heart rate, pupil diameter, muscle tension, electromyography (EMG), cortisol levels, galvanic skin response (GSR), respiration, and blood volume pulse (BVP), and video (facial expressions and gestures), (Kiesler, Zubrow, Moses, & Geller, 1985; Kramer, 1991; Wiethoff, Arnold, & Houwing, 1991; Scheirer, Fernandez, Klein, & Picard, 2002; Octavia, Raymaekers, & Coninx, 2011).

Human emotions can be characterized by arousal and valence. Arousal refers to emotional excitedness or activation, and ranges from calming or soothing to exciting or agitating. Valence refers to whether the emotional state of user is positive or negative, and ranges from highly positive to highly negative (Schlosberg, 1954; Russell, 1980; Frijda, 1986; Lang, Greenwald, Bradley, & Hamm, 1993; Kensinger, 2004). These two aspects of human emotion that can be measured (Lang, Greenwald, Bradley, & Hamm, 1993; Kim, Bang, & Kim, 2004; Nasoz, Alvarez, Lisetti, & Finkelstein, 2004; Li, & Chen, 2006). FaceReader automatically analyzes facial expressions to calculate arousal and valence. It detects ‘happy’ has mid-arousal and positive valence, while ‘angry’ has high arousal and negative valence (Loijens et al., 2012).

Facial expressions accompany muscle movements beneath the skin of the face. Since patterns of muscle activation can represents emotional states (Ekman, 1970), recording and analyzing an individual’s facial expression has been widely used to assess user emotional state such as anger, happiness, sadness, surprise, dislike, fear (De Silva, Miyasato, & Nakatsu, 1997; De Silva & Ng, 2000), and disgust (Benţa et al., 2009). For instance, facial electromyography (EMG) employs a sensor placed at the corrugator of the face to assess facial expression (Hazlett, 2003).

Skin conductance has been the basis of many sensor techniques, including galvanic skin response, electrodermal response, psychogalvanic reflex, skin conductance response, or skin conductance level (Conesa, 1995). For example, sweat gland activity raises skin conductance as a result of the sympathetic branch of the autonomic nervous system becoming highly aroused. Thus, GSR could be used for measuring emotional responses (Carlson, 2013).

Electrodermal activity (EDA) is defined as changes in the skin's electricity. An EDA sensor can detect autonomic changes in the electrical properties of the skin. Fluctuations and reduction in EDA signals were observed when the subjects were reminded of negative memories (Barrowcliff, Gray, Freeman, & MacCulloch, 2004). In contrast, higher EDA signals were detected when participants were shown evocative photos (Radin, 2004). The EDA sensor may be employed to investigate implicit emotional states, which may raise conscious awareness or intention of behavior (Braithwaite, Watson, Jones, & Rowe, 2013).

Delayed response and its impact in HRI are the focus of the study described in the nest section.

METHOD

Objective and Hypothesis

The objective of this study was to establish the effects of time delay on human emotion, physiological signals, and user satisfaction. Time delay of system is expected to results in increased negative emotions, increased arousal, and decreased user satisfaction for tasks of both low and high difficulty.

Participants

A total of 21 university students (14 males, 7 females), with an average age of 28.3 (range: 22 – 43) participated. All subjects were experienced computer users who had been using computers on average 3.8 hours in a day. Also, all participants had normal or corrected-to-normal vision in order to exclude the possibility of diminished attention due to vision problems.

Task

Participants navigated a robot in a simple and complicated (see Figure 1). Participants were spatially separated from the maze and could only see the video from the robot's camera.

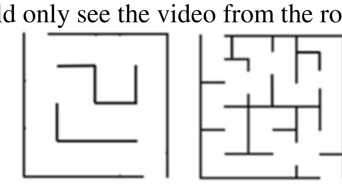

Figure 1. Silpe (left) and complicated (right) maze.

Independent Variables

The two independent variables were *Time Delay* (no time delay, time delay) and *Task Difficulty* (low, high). Time delay was manipulated via lags in the system feedback. Based on findings from a pilot study, the duration of each feedback delay varied randomly at either 2 or 3 seconds. Task difficulty was manipulated via maze complexity (see Figure 1).

Dependent Variables

The principal dependent variables were emotional state, physiological arousal, and user satisfaction. Emotional state had both a subjective component, reported via self-assessed emotion questionnaire, and an objective component, which utilized the FaceReader software. The subjective questionnaire of user emotion asked participants to rate, on a five-point Likert scale, seven different emotions – happy, angry, sad, surprised, scared, disgusted, and frustrated. To detect user arousal, an EDA sensor was used for detecting variation of physiological signals during the experiment. Finally, user satisfaction was measures with a subjective questionnaires. Table 1 shows the measurements of dependent variables.

Table 1. Dependent variables and associated metrics.

DV	Metric	Measure-ment	Frequency
Emotion	Questionnaire	Likert Scale 1 – 5	After each trial
	FaceReader	Continuous Scale 0 – 1	During each trial
Arousal	EDA	Microsiemens	During each trial
Satisfaction	Questionnaire	Likert Scale 1 – 5	After each trial

Experiment Design

This was a 2 (time delay: none vs. time delay) x 2 (task difficulty: low vs. high) within-subject design. Each condition was tested twice (eight trials). Trials were counterbalanced.

Procedure

The experiment began with consent, short briefing, demographic survey, and the attachment of physiological sensors. Participants were asked to read for 15 minutes while calibrating the EDA sensor. Participants were trained on robot operation. After every trial, participants completed a questionnaire on emotion and satisfaction. A post-experiment survey gathered participant opinions and strategies.

Testing Apparatus

Figure 2 shows the robot developed by Zhong (2013). The mazes measured 8.9 feet by 8.9 feet and were made of foam board. The EDA was placed around the wrist of the non-dominant hand of operators. The EDA sensor used was an Affectiva Q-sensor connected to a Dell Precision T1700 desktop PC. The robot measured 6 inch by 4 inch by 5.5 inch.

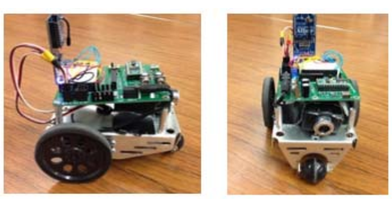

Figure 2. Robot's side and front view.

RESULTS

Shapiro-Wilk tests were used for normality; Bartlett's tests were used for homogeneity of variance. ANOVA results are reported as significant for alpha <.05.

Emotional State

Subjective Rating. Of the five emotions measured, only frustration and anger showed significant results. The participants' subjective rating of frustration showed that feedback delay in both simple (M=2.98, SD=1.25) and complicated maze (M=3.19, SD=1.29) was significantly ($F_{(1,19)}$=30.37, p<.0001) higher than no feedback delay of both simple (M=1.95, SD=0.89) and complicated maze (M=2.50, SD=1.06) (see Figure 3). The main effect of task difficulty on user frustration was also significant in time delay condition ($F_{(1,20)}$=6.33, p<.03). However, interaction between time delay and task difficulty was not significant.

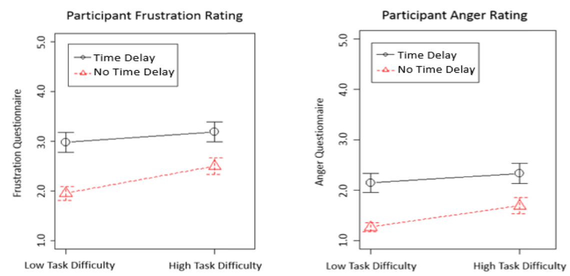

Figure 3. Frustration (left) and anger (right) results.

Anger showed a significant difference ($F_{(1,19)}$=29.179, p<.0001) with time delay, in both low (M=2.15, SD=1.22) and high task difficulty (M=2.33, SD=1.26) causing higher anger than the no time delay condition with both low (M=1.27, SD=0.55) and high task difficulty (M=1.69, SD=1.05) (see Figure 3). However, task difficulty and interaction between time delay and task difficulty were not significant.

Objective Results. Of the seven emotions classified by FaceReader, only anger was significantly ($F_{(1,20)}$=5.13, p<.05) affected by time delay (see Figure 4). The average intensity value of anger in time delay condition with both low (M=0.53, SD=0.24) and high task difficulty (M=0.57, SD=0.29) was higher than no time delay with both low (M=0.41, SD=0.25) and high task difficulty (M=0.46, SD=0.24). However, the task difficulty and interaction between time delay and the task difficulty were not significant.

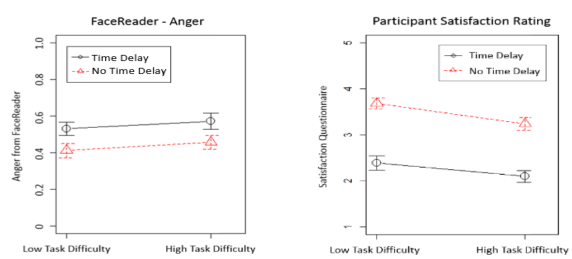

Figure 4. Anger (left) and user satsifatcion (right) results.

User Satisfaction

Time delay in both simple (M=2.39, SD=0.97) and complicated maze (M=2.09, SD=0.82) was significantly ($F_{(1,19)}$=98.37, p<.0001) lower than no feedback delay of both simple (M=3.68, SD=0.72) and complicated maze (M=3.24, SD=0.88) (Figure 4). The main effect of task difficulty on user satisfaction was also significant ($F_{(1,20)}$=12.37, p<.003) in both time delay and no time delay conditions. However, interaction between time delay and task difficulty was not significant.

Physiological signals: Electodermal Activity

Due to a loss of five participants' data, only 16 of the 21 participants' sensor data was used. EDA signal data were normalized to the participant's baseline data and averaged to create profiles of signals in different conditions. Figure 5 illustrates the time course of the (averaged) normalized EDA for the four conditions. The EDA signal collected showed a marked increase in the time delay over the no time delay condition, across both task difficulty conditions.

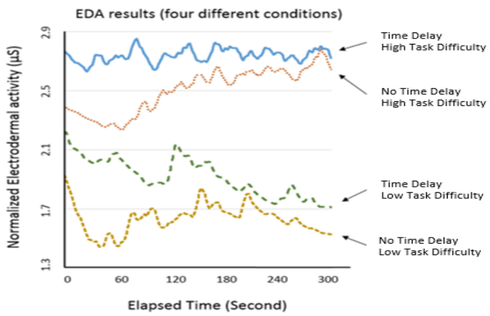

Figure 5. Average normalized change of electrodermal activity.

DISCUSSION

The results of the study show that time delay significantly influences emotional state (specifically frustration and anger), physiological arousal, and user satisfaction in robotic control task. These results held true for both low and high task difficulties. In the post-experiment questionnaires, participants reported that they felt higher frustration, anger with the time delay than no delay condition.

Among the seven emotions, only frustration and anger show significant relation to time delay. Frustration and anger have been shown to be related. Frustration is highly related to aggression (Dollard, Miller, Doob, Mowrer, & Sears, 1939; Miller, 1941; Morlan, 1949) and anger is regarded to contain higher emotional activation than frustration (Barrett, 2006; Lindquist, & Barrett, 2008). Comments from the participants included: "It is annoying," "I don't understand why it is not moving properly," "I'm almost angry."

Averaged results of normalized EDA showed different levels of arousal for four different combinations of time delay and task difficulty levels. In spite of steady occurrence of the feedback delay events throughout the trials, the EDA signals gradually decreased in low task difficulty condition after the middle of the trials. It is possible that the participants acclimatized to time delay during the trials. If this were the case, this has implication for the ability of EDA to detect arousal during long term exposure. This is an area where further work might establish relation between EDA and emotional excitedness or activation over long periods of time.

Successful systems support users to help maximize productivity and minimize overheads. According to DeLone and McLean (2002), one of the key measurements of system success in human-computer interaction is user satisfaction level. Likewise, operator satisfaction would be a key factor to design more desirable HRI. This study shows that operators were more dissatisfied with lags of system than without lags when they navigated mazes. Task difficulty was also influential factor to decrease user satisfaction.

These findings illustrate how time delay leads to negative emotions and lower user satisfaction. A deeper understanding of the relationship between system delays and the emotional repose can be used to inform the design affect-aware systems that adapt based on the user's emotional state. Together with the ability to detect user emotion, adaptive systems (Scerbo, Freeman, Mikulka, 2003; Feigh, Dorneich, & Hayes, 2012), could adjust its behavior to respond to negative emotions, may have the potential to enhance performance and productivity in the context of human-robot interaction.

REFERENCES

Adelstein, B. D., Lee, T. G., & Ellis, S. R. (2003, October). Head tracking latency in virtual environments: psychophysics and a model. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 47, No. 20, pp. 2083-2087). SAGE Publications.

Barber, R. E., & Lucas Jr, H. C. (1983). System response time operator productivity, and job satisfaction. Communications of the ACM, 26(11), 972-986.

Barrett, L. F. (2006). Valence is a basic building block of emotional life. Journal of Research in Personality, 40(1), 35-55.

Barrowcliff, A. L., Gray, N. S., Freeman, T. C., & MacCulloch, M. J. (2004). Eye-movements reduce the vividness, emotional valence and electrodermal arousal associated with negative autobiographical memories. Journal of Forensic Psychiatry & Psychology, 15(2), 325-345.

Benţa, K. I., Van Kuilenburg, H., Eligio, U. X., Den Uyl, M., Cremene, M., Hoszu, A., & Creţ, O. (2009). Evaluation of a System for RealTime Valence Assessment of Spontaneous Facial Expressions. In Distributed Environments Adaptability, Semantics and Security Issues International Romanian-French Workshop, Cluj-Napoca, Romania (pp. 17-18).

Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2013). A Guide for Analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for Psychological Experiments. Psychophysiology, 49, 1017-1034.

Carlson, N. R. (2013). Physiology of behavior. Pearson.

Casals, A. (1998). Robots in surgery. In Autonomous Robotic Systems (pp. 222-234). Springer London.

Conesa, J. (1995). Electrodermal palmar asymmetry and nostril dominance. Perceptual and motor skills, 80(1), 211-216.

Corde Lane, J., Carignan, C. R., Sullivan, B. R., Akin, D. L., Hunt, T., & Cohen, R. (2002). Effects of time delay on telerobotic control of neutral buoyancy vehicles. In Robotics and Automation, 2002. Proceedings. ICRA'02. IEEE International Conference on (Vol. 3, pp. 2874-2879). IEEE.

Davis, E. S., & Hantula, D. A. (2001). The effects of download delay on performance and end-user satisfaction in an Internet tutorial. Computers in Human Behavior, 17(3), 249-268.

DeLone, W. H., & McLean, E. R. (2002, January). Information systems success revisited. In System Sciences, 2002. HICSS. Proceedings of the 35th Annual Hawaii International Conference on (pp. 2966-2976). IEEE.

De Silva, L. C., Miyasato, T., & Nakatsu, R. (1997). Facial emotion recognition using multi-modal information. In Information, Communications and Signal Processing, 1997. ICICS., Proceedings of 1997 International Conference on (Vol. 1, pp. 397-401). IEEE.

De Silva, L. C., & Ng, P. C. (2000). Bimodal emotion recognition. In Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on (pp. 332-335). IEEE.

Den Uyl, M.J.; Van Kuilenburg, H. (2008). The FaceReader: Online Facial Expression Recognition. Proceedings of Measuring Behavior 2005, Wageningen, The Netherlands, August 30 - September 2, 2008, pp. 589-590.

Dollard, J., Miller, N. E., Doob, L. W., Mowrer, O. H., & Sears, R. R. (1939). Frustration and aggression. Yale University Press, New Haven

Dorneich, M. C., Ververs, P. M., Mathan, S., Whitlow, S., & Hayes, C. C. (2012). Considering etiquette in the design of an adaptive system. Journal of Cognitive Engineering and Decision Making, 6(2), 243-265.

Draper, J. V. (1993, December). Human factors in telemanipulation: perspectives from the Oak Ridge National Laboratory experience. In Optical Tools for Manufacturing and Advanced Automation (pp. 162-174). International Society for Optics and Photonics.

Drascic, D., Milgram, P., & Grodski, J. (1989, November). Learning effects in telemanipulation with monoscopic versus stereoscopic remote viewing. In Systems, Man and Cybernetics, 1989. Conference Proceedings, IEEE International Conference on (pp. 1244-1249). IEEE.

Ekman, P. (1970). Universal facial expressions of emotion. California Mental Health Research Digest, 8, 151-158.
Ellis, S. R., Mania, K., Adelstein, B. D., & Hill, M. I. (2004, September). Generalizability of latency detection in a variety of virtual environments. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 48, No. 23, pp. 2632-2636). SAGE Publications.
Feigh, K.M., Dorneich, M.C., & Hayes, C.C. (2012). "Towards a Characterization of Adaptive Systems: a Framework for Researchers and System Designers", Journal of Human Factors and Ergonomics. Vol. 54, No. 6, pp. 1008-1024.
Frijda, N. H. (1986). The emotions. Cambridge University Press.
Guynes, J. L. (1988). Impact of system response time on state anxiety. Communications of the ACM, 31(3), 342-347.
Hayes, C. C. & Miller, C. A. (2010). Should computers be polite? In C. C. Hayes & C. A. Miller (Eds.), Human-computer etiquette: Cultural expectations and the design implications they place on computers and technology (pp. 1–12). Boca Raton, FL: Taylor & Francis
Hazlett, R. (2003). Measurement of user frustration: a biologic approach. CHI'03 extended abstracts on Human factors (pp. 734-735). ACM.
Held, R., Efstathiou, A., & Greene, M. (1966). Adaptation to displaced and delayed visual feedback from the hand. Journal of Experimental Psychology, 72(6), 887.
Hone, K. (2006). Empathic agents to reduce user frustration: The effects of varying agent characteristics. Interacting with Computers, 18(2), 227-245.
Hoxmeier, J. A., & DiCesare, C. (2000). System response time and user satisfaction: An experimental study of browser-based applications. AMCIS 2000 Proceedings, 347.
Kensinger, E. A. (2004). Remembering emotional experiences: The contribution of valence and arousal. Reviews in the Neurosciences, 15(4), 241-252.
Kiesler, S., Zubrow, D., Moses, A.M., & Geller, V. (1985). Affect in computer-mediated communication: An experiment in synchronous terminal-to-terminal discussion. Human-Computer Interaction 1(1), 77-104.
Kim, K. H., Bang, S. W., & Kim, S. R. (2004). Emotion recognition system using short-term monitoring of physiological signals. Medical and biological engineering and computing, 42(3), 419-427.
Klein, J., Moon, Y., & Picard, R. W. (2002). This computer responds to user frustration: Theory, design, and results. Interacting with computers, 14(2), 119-140.
Klein, J., Moon, Y., and Picard, R. W. (1999). This computer responds to user frustration. In CHI'99 extended abstracts on Human factors in computing systems (pp. 242-243). ACM.
Klein, T., Rosalind, C., & Picard, W. (1999). Computer response to user frustration.
Kramer, A. F. (1991). Physiological metrics of mental workload: A review of recent progress. Multiple-task performance, 279-328.
Kuhmann, W., Boucsein, W., Schaefer, F., & Alexander, J. (1987). Experimental investigation of psychophysiological stress-reactions induced by different system response times in human-computer interaction. Ergonomics, 30(6), 933-943. Kuhmann, W. (1989). Experimental investigation of stress-inducing properties of system response times. Ergonomics, 32(3), 271-280.
Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. Psychophysiology, 30(3), 261-273.
Li, L., & Chen, J. H. (2006). Emotion recognition using physiological signals. In Advances in Artificial Reality and Tele-Existence (pp. 437-446). Springer Berlin Heidelberg.
Lindquist, K. A., & Barrett, L. F. (2008). Emotional complexity. Handbook of emotions, 513-530.
Loijens, L., Krips, O., Van Kuilenbug, H., Den Uyl, M., Ivan, P., Theuws, H., and Spink, A. (2012) FaceReader Reference Manual Version 5.0, Noldus Information Technology b.v.
Luck, J. P., McDermott, P. L., Allender, L., & Russell, D. C. (2006, March). An investigation of real world control of robotic assets under communication latency. In Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction (pp. 202-209). ACM.
Lum, M. J., Rosen, J., Lendvay, T. S., Sinanan, M. N., & Hannaford, B. (2009, May). Effect of time delay on telesurgical performance. In Robotics and Automation, 2009. ICRA'09. IEEE International Conference on (pp. 4246-4252). IEEE.
Miller, C.A. & Dorneich, M.C. (2006), "From Associate Systems to Augmented Cognition: 25 Years of User Adaptation in High Criticality Systems", Proceedings of the 2nd Augmented Cognition International, San Francisco, CA.
Miller, C. A., & Parasuraman, R. (2007). Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control. Human Factors: The Journal of the Human Factors and Ergonomics Society, 49(1), 57-75.
Miller, N. E. (1941). I. The frustration-aggression hypothesis. Psychological Review, 48(4), 337.
Morlan, G. K. (1949). A note on the frustration-aggression theories of Dollard and his associates. Psychological review, 56(1), 1.
Nasoz, F., Alvarez, K., Lisetti, C. L., & Finkelstein, N. (2004). Emotion recognition from physiological signals using wireless sensors for presence technologies. Cognition, Technology & Work, 6(1), 4-14.
Nonami, K., Shimoi, N., Huang, Q. J., Komizo, D., & Uchida, H. (2000). Development of teleoperated six-legged walking robot for mine detection and mapping of mine field. In Intelligent Robots and Systems, 2000. (IROS 2000). Proceedings. 2000 IEEE/RSJ International Conference on (Vol. 1, pp. 775-779). IEEE.
Octavia, J. R., Raymaekers, C., & Coninx, K. (2011). Adaptation in virtual environments: conceptual framework and user models. Multimedia Tools and Applications, 54(1), 121-142.
Owen-Hill, A., Suárez-Ruiz, F., Ferre, M., & Aracil, R. (2014, January). Effect of Video Quality and Buffering Delay on Telemanipulation Performance. In ROBOT2013: First Iberian Robotics Conference (pp. 555-568). Springer International Publishing.
Picard, R. W., & Klein, J. (2002). Computers that recognize and respond to user emotion: theoretical and practical implications. Interacting with computers, 14(2), 141-169.
Powers, S.., Rauh, C., Henning, R.A., Buck, R.W., & West, T.V. (2011). The effect of video feedback delay on frustration and emotion communication accuracy. Computers in Human Behavior, 27(5), 1651-1657.
Prewett, M. S., Johnson, R. C., Saboe, K. N., Elliott, L. R., & Coovert, M. D. (2010). Managing workload in human–robot interaction: A review of empirical studies. Computers in Human Behavior, 26(5), 840-856.
Radin, D. I. (2004). Electrodermal presentiments of future emotions. Journal of Scientific Exploration, 18(2), 253-273.
Ridao, P., Carreras, M., Hernandez, E., & Palomeras, N. (2007). Underwater telerobotics for collaborative research. In Advances in Telerobotics (pp. 347-359). Springer Berlin Heidelberg.
Rowe, D. W., Sibert, J., & Irwin, D. (1998). Heart rate variability: Indicator of user state as an aid to human-computer interaction. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 480-487). ACM Press/Addison-Wesley Publishing Co.
Russell, J. A. (1980). A circumplex model of affect. Journal of personality and social psychology, 39(6), 1161.
Scerbo, M. W.; Freeman, F. G. & Mikulka, P. J. (2003). "A brain-based system for adaptive automation". Theoretical Issues in Ergonomic Science, 4, 200-219
Scheirer, J., Fernandez, R., Klein, J., & Picard, R. W. (2002). Frustrating the user on purpose: a step toward building an affective computer. Interacting with computers, 14(2), 93-118.
Schleifer, L. M., & Amick III, B. C. (1989). System response time and method of pay: Stress effects in computer-based tasks. International Journal of Human-Computer Interaction, 1(1), 23-39.
Schlosberg, H. (1954). Three dimensions of emotion. Psychological review, 61(2), 81.
Sears, R. R. (1937). Initiation of the repression sequence by experienced failure. Journal of Experimental Psychology, 20(6), 570.
Sheik-Nainar, M. A., Kaber, D. B., & Chow, M. Y. (2005). Control gain adaptation in virtual reality mediated human–telerobot interaction. Human Factors and Ergonomics in Manufacturing & Service Industries, 15(3), 259-274.
Shneiderman, B. (1992). Designing the user interface: strategies for effective human-computer interaction (Vol. 2). Reading, MA: Addison-Wesley.
Szameitat, A. J., Rummel, J., Szameitat, D. P., & Sterr, A. (2009). Behavioral and emotional consequences of brief delays in human–computer interaction. International Journal of Human-Computer Studies, 67(7), 561-570.
Wiethoff, M. A. A. G., Arnold, A. G., & Houwing, E. M. (1991). The value of psychophysiological measures in human-computer interaction. Human-Aspects in Computing: Design and Use of Interactive Systems and Work with Terminals. Amsterdam: Elsevier, 661-665.
Zhong, P. (2013). Perception in Remote Navigation. Ph.D. Dissertation, Iowa State University.

265

Introducing Change Into Complex Systems

Introducing Change into Complex Systems

Panel Organizer
Katherine P. Kaste
Naval Air Warfare Center Training Systems Division (NAWCTSD)
Orlando, FL

Panelists
Dr. Robert Hoffman, Institute for Human and Machine Cognition
Dr. Barrett Caldwell, Purdue University
Dr. Nicholas Kasdaglis, U.S. Army Research Laboratory
Dr. Kelly J. Neville, Naval Air Warfare Center Training Systems Division (NAWCTSD)

Change in a complex system—for example, to its technology, procedures, or information flows—no matter how small, has the potential to create large effects and ripples of disruption. A complex system's dynamics cannot be fully known, and the effects and disruptions produced by change are difficult to predict. Nonetheless, complex systems can be at least partly understood in terms of patterns in their dynamics, generalizable principles, and mechanisms of control, balance, and resilience. This panel will focus on complex-systems theory, research and what they suggest about how to introduce change into a complex system such that system resilience and health are disrupted minimally. Case studies may be discussed, as well; examples of changes to established complex systems. These changes include introducing remotely piloted aircraft systems (RPAS) into the National Airspace System (NAS), additional automation into air traffic control, and new technology into military air combat training.

COMPLEXITY THEORY

A complex system is a system consisting of many interacting parts. The interactions are not specified, like a clock's inner workings. Instead, complex systems are challenging to work with and understand as they are neither "rigidly ordered nor rigidly disordered" (Arrow, McGrath, Berdahl, 2000, p.38). This means that although their behavior is not completely predictable, nor is it completely unpredictable. Their behavior tends to conform, if only loosely, to emergent patterns of behavior that support resilience (i.e., the ability of the system to withstand unusual events and external pressures).

Because humans are complex systems unto themselves, any system in which they participate becomes an increasingly complex system. The behavior of these systems becomes less predictable and more challenging to understand and control. These systems encompass a wide variety of work teams (teams of humans as well as teams of technology and humans) and organizations, including power plant operations, flight operations, air traffic control, transportation systems, the operation of individual vehicles, emergency response operations, and much more.

Because they are not entirely predictable and can have many moving and interacting parts, designing, redesigning, and introducing change into a complex system is especially challenging. Yet, as Human Factors professionals, these are the types of systems we are trying to improve.

This panel aims to explain what a complex system is, share examples of current complex systems and provide suggestions on how to introduce change into complex systems without causing catastrophic disruptions. It will additionally explore the adequacy of current methods and approaches of the human factors and systems engineering communities for handling complex systems problems and will discuss new methods, concepts, and approaches that could assist practitioners in these communities as they increasingly confront complex systems problems.

The requirements to be on the panel included experience studying or working with complex systems and current involvement in research and or theories related to the study, design, and improvement of complex systems. The panelists' work covers a range of problem domains and they represent a variety of perspectives as they come from both government research labs and academia.

DR. ROBERT HOFFMAN

Taking "Systems" Seriously: Principles and Measures

If we reserve the word "system" to refer to collectives of humans and machines that conduct macrocognitive work, then certain powerful principles of complexity are necessarily entrained. Examples are the principle of irreducible complexity, the principle of necessary incompleteness, the principle of indeterminate consistency, and the principle of emergence. As I will explain in my presentation, these aspects of systems thinking seem counterintuitive and at odds with traditional theoretics and methodology, which emphasize reduction over elegance. At the same time that complexity holds sway, there are also fundamental bounds on macrocognitive work systems. These are manifestations of indeterminate causality: bounded ecology, bounded cognizance, bounded perspectivity, bounded responsibility, and bounded effectivity. Each of the bounds entails inescapable trade-offs that must be considered in the design and analysis of macrocognitive work systems. As trade-off relations, each references two crucial conceptual measurables. While it is a challenge to develop operational definitions of the conceptual measurables, such as "adaptivity" and "resilience," it is not impossible, and progress is being made along these lines.

DR. BARRET CALDWELL

In order to effectively address the problem of unintended consequences in complex systems, we must often face the challenges inherent in only experiencing "the arrow of time" in one direction, and the essential characteristic of experience as what one gains after one needs it. The study of individual brain function shows remarkable ranges of complexity, plasticity, and adaptability, highlighting the human ability to both create unique experiences and associations, and respond to a vast array of inputs and generated thoughts. The very elements of modern information technology and continuous innovation pathways allow for integration of technologies and thoughts and goals in novel ways, at daunting levels of mathematical combinations. Thus, the concept of a simple clockwork organization of human-scale systems is clearly and essentially inadequate to describe and analyze the very entities that we hope to create and control.

A longstanding thread of research in human-machine interaction has shown that human operators have difficulty in managing supervisory or operational control of poorly coupled, feedback delayed, or otherwise "noisy" systems. In addition, cognitive ergonomics and cognitive science studies demonstrate that many people (including experts) can have substantial challenges in effectively anchoring probabilities and large scale, second- and higher-order effects that complicate cause and effect relationships. In essence, people have trouble with nonlinear and uncertain system dynamics, particularly those that can be described as incorporating elements of "deep uncertainty". However, it is exactly these issues that are the most distinctive and representative characteristics of complex systems, especially those that operate in cyber-physical, sociotechnical, and other hybrid contexts.

In this talk, I will discuss prior research and past work in addressing operational stability ranges, including resilience and robustness of complex dynamic systems. I will also discuss cognitive, social, and operational limits to obtaining fresh, accurate, and actionable information to support effective decisions (and more importantly, performance actions) in distributed and coordinated task environments. Examples from natural weather event response, aerospace operations, and healthcare environments are discussed from a combination of mathematical, organizational, and sociotechnical perspectives.

DR. NICHOLAS KASDAGLIS

Complex socio-technical systems exhibit emergent behavior. That is, as a system's dimensions and interrelations increase, unique, and often unanticipated system behaviors and properties emerge. Emergent behavior has been identified in notable accidents of the 20th century (Perrow, 1984). MIT scientist Nancy Leveson's (2004) sobering warning is appropriate in this context "we are attempting to build systems that are beyond our ability to intellectually manage" (p.2). Although some researchers have considered the implications to safety and efficiency of future systems, more attention must be given. Often Human Factors researchers seek to anticipate real world system behaviors by inference derived from carefully controlled experimental research. The result has at times been unfortunate; in fact, the oft cited term automation surprisingly (Sarter, Woods, & Billings, 1997) appears to be the "normal" behavior of complex systems operating in a dynamic world. Incremental changes to a system's context of use, and technology can introduce such surprises. Thus, some researchers have called for revision to existing paradigms of modeling and evaluation of complex socio-technical systems (Boy, 1998; Hollnagel, 2012); accident causation (Kasdaglis & Oppold 2014); safety (Leveson, 2004); situation awareness (Kasdaglis, Newton, & Lakhmani, 2014), and the general understanding of joint cognitive systems (Hollnagel & Woods, 2005).

Complex socio-technological systems, such as commercial aviation, while offering highly reliable technology, have not necessarily met the challenge of emergent system behaviors that result from dysfunctional interactions between human and automation. In-Flight Loss of Control (I-LOC) events, the leading cause of commercial aviation fatalities, are one class of accidents that have resulted from such emergent behavior. For an example, one needs to look no further than the 2009 Air France crash of an A330. This highly automated aircraft reverted to a mode of operation that befuddled the pilots to the extent that the aircraft plummeted into the ocean from cruising altitude. My present research at the Human Centered Design Institute of Florida Institute of Technology is attempting to address such undesirable interactions between human and automation in complex systems.

DR. KELLY J. NEVILLE

The U.S. Navy is considering the use of a new training paradigm and associated technology in air combat training, which means they will be introducing change into an established complex system—Navy air combat training. It is difficult to anticipate the effects of change within a complex system and, in this particular system, the consequences of a malfunction or error can be grave. Consequently, the U.S. Navy initiated a research program to attempt to anticipate and mitigate potential negative effects of *Live, Virtual, and Constructive (LVC) training* on air combat training. Within this program, we conducted a series of semi-structured interviews with active and former air combat personnel to understand ways LVC training technology might contribute to the development of a potential hazard situation in live air combat training and ways those hazard situations might be mitigated through either design or procedural changes.

After identifying a number of potential hazard situations, we asked two air combat subject matter experts (SMEs) to independently vet and help us better to understand them. In the case of almost all identified hazard situations, the SMEs were able to identify one or more mechanisms within the existing training system that would prevent the situation from occurring. For example, the replacement of live adversary pilots with computer-generated constructive adversaries would mean that adversary pilots would not be able to detect training rule violations by friendly-force pilots (i.e., by trainees). This would be a weakness in the training system and would pose a threat to both safety and training, but it is mitigated by redundancy in the form of other exercise participants and personnel, such as the pilot's wingman and the Range Training Officer (RTO), who share the responsibility of detecting training rule violations. This is also an example of flexibility within the system; there is flexibility in that multiple means exist for accomplishing the task of monitoring training rule adherence.

Flexibility and redundancy, especially functional redundancy where the resources that can accomplish a given task are not identical to one another and so are not vulnerable to the same threats, are two characteristics that give resilience to complex systems. Furthermore, they are examples of characteristics that tend to be common to resilient complex systems across domains (e.g., Bar-Yam, 2004; Kitano, 2004).

We found that we were able to align each hazard-avoidance mechanism the SMEs identified with a general characteristic of resilient complex systems. Our analysis consequently evolved into one of assessing the potential hazard situations in terms of complex system resilience mechanisms. Each hazard is now being assessed in terms of whether it (a) poses a threat to a mechanism the training system uses to maintain its resilience, such as flexibility or redundancy, (b) is not mitigated by an existing resilience mechanism, or (c) threatens a core system goal, which, in the case of air combat training, is to develop a level of proficiency and combat preparedness beyond what can be gained using simulators.

This approach to assessing the effects of change in a complex established system has at least a couple of benefits. As one example, it facilitates the identification of problems and solutions that relate to the general resilience of the training system over those that address one system element at a time and may have unanticipated interactions with other elements. In addition, the approach is grounded in complex systems principles that have been identified as core to the health of complex systems across a variety of domains, lending greater credibility to our findings and recommendations.

DISCUSSION QUESTIONS:

- During your professional experience, have you encountered difficulties in introducing change to complex systems? If so, please explain.
- Can you describe results of your complexity research?
- What are examples of real-world complex systems failing?
- In the work you described, what is the impact of feedback in complex systems?
- With the continuous flow of information into the system, are there examples where thresholds are being approached and the system auto adjusts in order to process, even when the human element is not actively monitoring all available feedback?
- How can we address these issues with research? What are the implications for system acquisition?
- Are there fundamental differences between natural systems maintaining balance and health versus engineered systems' ability to maintain health and balance?
- Do you believe there is a role for automation in complex systems? How do you account for the uncertainty involved?
- Can automation "handle" complex systems?
- Are there implications to expert/novice differences in resilience engineering?
- Are decision support tools able to aid operators in complex systems?
- Do you foresee empirical evaluations important to investigate areas discussed?

AUTHOR'S NOTE

The views expressed herein are those of the authors and do not necessarily reflect the official position of the Department of Defense, its components, or the organizations with which the individuals are affiliated.

REFERENCES

Arrow, H., McGrath, J., E., & Berdahl, J. L. (2000). *Small Groups As Complex Systems: Formation, Coordination, Development, and Adaptation.* Thousand Oaks, CA: Sage Publications, Inc.

Bar-Yam, Y. (2004). *Making things work*. Boston, MA: NECSI Knowledge Press.

Boy, G. A. (1998). *Cognitive function analysis* Stamford, CT: Ablex Publishing Corporation.

Hollnagel, E. (2012). *Fram: The functional resonance analysis method: Modelling complex socio-technical systems.* England: Ashgate Publishing, Ltd.

Hollnagel, E., & Woods, D. D. (2005). *Joint cognitive systems: Foundations of cognitive systems engineering*. Boca Raton, FL: CRC Press.

Kasdaglis N., Oppold. P. (2014). Surprise, attraction, and propagation: An aircraft is no place for a catastrophe.

Proceedings of the Human Factors and Ergonomics Society58th Annual Meeting [CD-ROM].

Kasdaglis, N., Newton, O., & Lakhmani, S. (2014). System state awareness: A human centered design approach to awareness in a complex world. Paper presented at the *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* [CD-ROM].

Kitano, H. (2004). Biological robustness. *Nature Review Genetics*, *5*, 826-837.

Leveson, N. (2004). A new accident model for engineering safer systems. *Safety Science, 42*(4), 237-270.

Perrow, C. (1984). *Normal accidents: Living with high-risk technologies*. New York: Basic Books.

Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). Automation surprises. *Handbook of Human Factors and Ergonomics, 2*, 1926-1943.

270

Addressing Complex Decision-Making for Unmanned Aerial System Sensor Operators

Addressing Complex Decision-Making for Unmanned Aerial System Sensor Operators

The purpose of this presentation is to outline a current Office of Naval Research (ONR) Basic Research Challenge (BRC) effort investigating the information complexity Unmanned Aerial Systems (UASs) Sensor Operators (SOs) must endure in order to effectively make decisions. UASs are increasingly becoming critical systems within the Department of Defense as their capabilities have been deemed highly successful. However, certain aspects of UASs cause their operations to be highly complex for the human in the loop. Not only do they lack the physical sensations from the aircraft that manned operators receive, but they must translate mass amounts of disparate data into actionable information. Furthermore, current operations require a team of individuals to operate a UAS, however as technology develops there is an emerging requirement for one operator to control multiple systems, or adopt a one-to-many approach. This poster will present the support and theories behind the effort to address these issues and will discuss future application.

INTRODUCTION

A current challenge within the unmanned community is the ability to assimilate large amounts of data into actionable information. This challenge may become even more aggravated as the requirement for one-to-many operation becomes increasingly salient. This requirement will cause the attention of operators to become heavily stressed.

Currently, this challenge is being managed by experienced SOs who are responsible for controlling the sensors and cameras to aid the identification and tracking of areas of interest. Not only do SOs have a large amount of information to receive, interpret, and respond to, the information may be incomplete or lacking in detail, further complicating their tasks.

In order to address these issues, the BRC effort is investigating how to enhance SO performance through system design centered on human cognition and decision making theories. One way to mitigate the challenges they face is to aid their performance through interface design.

This effort aims to produce display design guidelines that will enable novices to act as experts when faced with massive amounts of information in highly complex settings. These guidelines will be supported by empirical evaluation and basic human factors principles to be generalizable across UAS platforms.

BACKGROUND

Decision-Making

Imperative to this research is an understanding of the cognitive decision-making techniques experts use in multiple types of environments. Cognitive decision-making has been a focus area for decades across multiple domains to include master chess matches and landmine detection (Chabris & Hearst, 2003; Lesgold et al., 1998). Through this research, multiple techniques have been identified that experts may engage in when processing information. These techniques are not used exclusively; rather they seem to be used in combinations to effectively complete tasks.

The Chunking (Miller, 1956) and Template Theories (Gobet & Clarkson, 2004; Gobet & Simon, 1996) state experts rely on "chunking" items into familiar units for easier recall because we are limited by our memory capabilities. The two theories differ in exactly how experts use chunks – whether they are able to hold large chunks within their memory or if they hold multiple smaller chunks in linked "chains." At the root of this research is the debate regarding the size of chunks and the limitations of human memory (Gobet & Clarkson, 2004).

Experts also employ the use of schemas, often referred to as "scripts" when recalling information or making assumptions based upon the information presented. Schemas are generalizations of how we perceive the world around us, based upon our past experience (Bartlett, 1932 cited from Roediger, 2000). Thus they differ as they are dependent upon the amount of experience of an individual.

Radiology research states experts have the ability to adjust and modify their schemas based upon incoming information. Novices tend to make quick decisions once they believe the information fits their schema, no matter if other information negates the original conclusion. Experts can effectively explain away erroneous data as the schemas help *guide* their search, and does not force the direction of their decisions (Lesgold et al., 1988). Novice radiologists make mistakes as they jump to the first diagnosis they arrive at after comparing shapes seen within x-rays to those they had seen in textbooks. Experts, from experience, exhibited behaviors such as discussing what they were thinking out loud, explaining possibilities and referencing past cases to arrive at their decisions.

Bartlett's (Bartlett, 1932 cited from Roediger, 2000) "War on Ghosts" research also illustrates how our own schemas can lead us to conclusions that are not always true. While the conclusions we arrive at may make contextual sense based upon our experience, they may not make sense based on the present context at hand. Schemas are useful, however, as they provide a basis to refine incoming information.

Chess literature has also provided insight into memory limitations and expert ability to excel at a complex game. The literature states the use of pattern recognition or search-ahead techniques (Chabris & Hearst, 2003). Here, grand master level players placed under an immense time pressure condition are able to recognize patterns within the chess pieces to respond accordingly and force their opponents into less than desired positions, eventually winning. If they are not under time pressure, experts can search ahead and mentally visualize possible moves and outcomes to make the best decision possible. This research highlights the expert ability to remember moves and anticipate possible moves and outcomes of their opponents.

Finally, "frames" or mental models are structures used by experts to "account for data" (Klein, Phillips, Rall & Peluso, 2007). The sensemaking model outlines how individuals make sense of the world around them by ingesting and explaining away necessary information. Individuals actively react to their environment based upon cues, or anchors, that are recognized within the information at hand. This model stresses the action of the decision maker as it outlines whether or not the operator is actively seeking information to affirm his schema, negating information, or reanalyzing information that may have been previously discarded.

This review of decision making techniques allows the researchers to understand how people in different environments handle information and make decisions. These techniques will provide the foundation for manipulation of experimental variables to test if SOs also engage in these techniques, a combination of these techniques, or if there are novel techniques not yet outlined.

Interface Design and Automation

Workload can be high in complex work settings and automating tasks may help alleviate these burdens. However, questions arise when automation is considered. What should be automated and at what levels of automation? Guidelines for automation have been developed for the aviation domain (see Billings, 1991), however, the tasks and requirements of UASs differ from manned vehicle tasks and requirements. Consequently, additional research is needed to develop automation guidelines for UAS operator tasks to determine what should be automated and at which levels (McCarley & Wickens, 2005).

While automation taxonomies of varying levels and spectrums (ranging from full automation to full manual control) have been proposed in the literature (Sheridan & Verplanck, 1978; Parasuraman, Sheridan, & Wickens, 2000), it is important to note that there are certain advantages and disadvantages associated with different levels of automation. Systems that are completely manual with no automation may increase workload. On the other end of the spectrum, fully automated systems may decrease workload but operators may become complacent and over rely on the system to perform the operator's tasks (Parasuraman, Malloy, & Singh, 1993). Others have argued that when considering automation the question should not be to automate or not, but rather the question should be how automation should be designed so that it takes into account the system environment and human capabilities and limitations (Fern et. al., 2011).

Many times, operators are overloaded with incoming visual and auditory information that they must keep in their working memory and compile with other information. The demands of these tasks increase the workload burden, and if the burden is beyond the recommended threshold, operators can make errors that may lead to catastrophic incidents.

Additionally, Wickens' (1976) Multiple Resource Theory (MRT) provides one explanation as to why operators become overloaded. MRT states that the completion of tasks requires specific resources (i.e., visual, auditory, and mental), however, there is a limited amount of resources available. Some tasks may require more resources than available, and when simultaneous tasks are conducted, these tasks would likely compete for the same limited resources (Wickens, 1984). When the amount of resources is inadequate, this creates a heavy workload burden on the operator. To help reduce the impact of limited resources, displays are used to aid users in understanding visual or auditory information. A well designed display can reduce workload by mimicking expert operator decision making strategies such as pattern recognition (Sanderson, Pipingas, Danieli, & Silberstien, 2003; Bennett & Flach, 1992). However, poorly designed displays can actually increase the operator's workload burden. Complex environments may require multiple displays as the required information may not fit on one display. Moreover, clumsy organization of information on the interface can add to the already complex work environment (Bennett & Flach, 2011).

Additionally, monitoring multiple windows or displays and transitioning between those displays may lead to high visual workload. Operators of complex interface designs can easily "get lost" in transitioning between windows such that they "do not know where they are, where they came from, or where they have to go next" (Ziefle & Bay, 2006, p. 395). In cases where the work environment requires multiple displays, Woods (1984) suggests that the interface should support visual momentum. This means the interface supports the "user's ability to extract and integrate information across displays" (Woods, 1984, p.231). If visual momentum is high, the interface supports transitions from one display to the next. In contrast, low visual momentum would result in poor transitions such that a new display the operator moves to appears to be independent from the previous displays (Woods, 1984). When designing displays for complex work environments, designers need to be concerned with not only representing the required information of that work environment but also the transitions between multiple displays.

Current SO Operations

Currently, UAS SOs are required to interact with multiple screens, team members and technology in order to complete their tasks. While they have flexibility in how they configure their displays, some of the information is not readily available such as differences in metrics used between team members, knowledge of aircraft behavior and health status due to system design (In-person Interview, July 2014). SOs have determined specific workarounds in order to complete their tasks, however workarounds are not always ideal. These workarounds identify areas where limitations are present within the current operations and thus need to be addressed. Observations from one UAS platform identified visual momentum issues for the SO. For this specific platform, the lack of momentum may increase the workload encountered by the SOs which consequently increases the likelihood of error and difficulty experienced. More investigation must be completed to be able to generalize these assertions to other UAS platforms.

NEW CONTRIBUTION

In order to lessen SO workload, basic experimentation needs to be conducted to investigate how to effectively combine large amounts of disparate data into a manageable format. However, due to the nature of SOs in the UAS domain, it is important to investigate multiple facets involved with human interaction and technology. Simply focusing on the human itself or the system alone would not accurately address the problem. Specifically, it is imperative to investigate the type of tasks SOs must complete, the type of stressors specific to the SO domain, the cognitive abilities of the current experts, the intricacies of the interface used, and failure points between human and system. To contribute to this problem area, our approach includes the combination of cognitive psychology, decision theory, and machine learning methods into a unified framework.

This approach lends a more holistic product as it approaches the problem from multiple areas of research. As stated previously, when designing new interfaces (or guidelines for interface designs) it is imperative that human capabilities and limitations are taken into account as to not negatively impact performance.

Through basic experimentation, consisting of three interrelated studies, this effort will provide guidance for system development that enhances operator decision making through automation, data fusion, and enhanced visualizations.

Qualitative and quantitative physiology-based metrics collected real time will be included for the workload analyses (e.g., Bedford Scale and Modified Cooper-Harper; EEG, eye tracking) to provide further support for our guidelines. The qualitative and quantitative workload analysis will be compared to determine the degree of correlation in order to validate our workload assessment.

CURRENT STATUS

To approach this problem from the Human Systems Integration perspective, the BRC effort plans to conduct three interrelated experiments:

- Experiment 1 objective is to determine SO expert decision making behavior in multiple levels of complexity (baseline)
- Experiment 2 will focus on the effects of data integration and automation allocation on operator performance
- Experiment 3 will determine if the novel approaches developed aid performance

The experiments, scheduled for later in 2015, will be performed in a UAS testbed that allows for experimental manipulation of the levels of complexity experienced by the operator.

In preparation for experiment 1, three interviews have been completed and are currently under analysis. The cognitive task analysis (CTA) provides a comprehensive understanding of the tasks SOs must complete and the decision-making methods they use. The results of the CTA will also provide the appropriate tasks for experimentation as they must be realistic in complexity in order to effectively investigate SO tasks. The CTA will also provide details as to where human and machine interaction fails and thus may benefit from a decision support tool (i.e., display). CTAs are used to provide an understanding of the human capabilities and limitations in the context in which automation should be used and how those capabilities and limitations will benefit from automation (Fern et al., 2011). Instead of using blanket-statements to decide if tasks should be automated or not, Fern and colleagues (2011) suggests that automation is context specific and is derived from CTAs to identify the context in which automation should be used. As a result, automation is designed such that it does not go unused and does not hinder the operator in performing his tasks.

The CTA, along with results from experiment 1, will feed into experiment 2 where tasks will be automated and integrated based upon human cognition and workload assessment (i.e., qualitative responses and physiological analysis). Further into the effort, experiment 3 will investigate if operator performance was affected due to novel integration techniques.

BENEFITS

The approach discussed aims to help guide interface development for future UASs. By aiding interface development using the results of basic experimentation implementing multiple areas of research, a more successful design may be obtainable. Multiple aspects of performance/capabilities need to be addressed when suggesting changes to complex systems. These guidelines have the potential to create systems that optimize situational awareness and workload thus improving performance.

Design and automation guidelines can be used for SO tasks and displays across platforms for UASs. Furthermore, these guidelines can be applied to prototypes for SOs to investigate which display is best for performance. Operators may be required to control multiple systems and these guidelines will provide a starting point as to how data can be integrated and displayed and how automation should be used. Future research is needed to evaluate how these guidelines can be implemented.

AUTHOR'S NOTE

The views expressed herein are those of the authors and do not necessarily reflect the official position of the Department of Defense, its components, or the organizations with which the individuals are affiliated.

REFERENCES

Bennett, K.B. & Flach, J.M. (2011). *Display and interface design: Subtle science, exact art.* Boca Raton, FL: CRC Press.

Bennett, K. B., & Flach, J. M. (1992). Graphical displays: Implications for divided attention, focused attention, and problem solving. *Human Factors, 34*, 513–533.

Billings, C. E. (1991). *Human-centered aircraft automation: A concept and guidelines* (NASA Technical Memorandum 103885). Moffet Field, CA: NASA-Ames Research Center.

Chabris, C. F., & Hearst, E. S. (2003). Visualization, pattern recognition, and forward search: effects of playing speed and sight of the position on grandmaster chess errors. *Cognitive Science*, *27*, 637-648.

Fern, L., Shively, R. J., Draper, M. H., Cooke, N. J., Oron-Gilad, T., & Miller, C. A. (2011). Human-automation challenges for the control of unmanned aerial systems. In *Proceedings of the Human Factors and Ergonomics Society 55th Annual Meeting* [CD-ROM].

Gobet, F., & Clarkson, G. (2004). Chunks in expert memory: Evidence for the magical number four ... or is it two? *Memory, 12* (6), 732- 747.

Gobet, F., & Simon, H. A. (1996). The roles of recognition processes and look-ahead search in time-constrained expert problem solving: evidence from grandmaster level chess. *Psychological Science*, 7, 52-55.

Klein, G., Phillips, J. K., Rall, E. L., & Peluso, D. (2007). A Data-Frame Theory of Sensemaking. In a Robert A. Hoffman (Ed.) *Expertise Out of Context: Proceedings of the Sixth International Conference on Naturalistic Decision Making* (pp. 113 – 153). New York, NY: Taylor & Francis Group, LLC.

Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a Complex Skill: Diagnosing X-Ray Pictures. In Chi, M., Glaser, R., & Farr, M. (Eds.), *The Nature of Expertise* (pp. 311 -342). Hillsdale, NJ: Erlbaum.

Miller, G.A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review*, 63, 81-97.

McCarley, J. S., & Wickens, C. D. (2005). *Human factors implications of UAVs in the national airspace* (Tech. Rep. AHFD-05–05/ FAA-05-1). Savoy, IL: University of Illinois, Aviation Human Factors Division.

Parasuraman, R., Molloy, R., & Singh , I. L . (1993). Performance consequences of automation-induced ‘‘complacency.’’ *International Journal of Aviation Psychology, 3*, 1 – 23.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transaction on Systems, Man and Cybernetics-Part A: Systems and Humans*, *30*(3). 286-297.

Roediger, H. L., Bergman, E. T., & Meade, M. L. (2000). Repeated reproduction from memory. *Bartlett, culture and cognition* (pp.115-134). England: Psychology Pres..

Sanderson, P., Pipingas, A., Danieli, F. & Silberstein, R. (2003). Process monitoring using a configural display: human performance, self-report, and neuroimaging. *Theoretical Issues in Ergonomic Science, 4,* 151–174.

Sheridan, T. B. & Verplank, W. L. (1978). *Human and Computer Control of Undersea Teleoperators*. Cambridge, MA: Tech Rep.

Wickens, C.D. (1976) The effects of divided attention on information processing in tracking. *Journal of Experimental Psychology: Human Perception and Performance,1*, 1-13.

Wickens, C. D. (1984). Processing resources in attention. In Parasuraman, R., & Davies, R. (Eds.), *Varieties of Attention,* (pp. 63-101). New York: Academic Press.

Woods, D. D. (1984). Visual momentum: A concept to improve the cognitive coupling of person and computer. *International Journal of Man–Machine Studies, 21,* 229–244.

Ziefle, M., &. Bay. S. (2006). How to overcome disorientation in mobile phone menus: A comparison of two different types of navigation aids. *Human–Computer Interaction, 21*, 393–433.

446

Recognizing the Role of Consistency in a Delayed Memory Task

Recognizing the Role of Consistency in a Delayed Memory Task

Past research on short-term memory decay has found that participants are more efficient at remembering information when the delay between stimuli presentation and recall is short as opposed to long. In the current study we used Potential Performance Theory (PPT) to identify the role that both random and systematic factors play in observed memory performance over a delay. We presented participants with a string of letters followed by either a 2-second or 16-second delay. Following the delay, participants were presented with a two alternative forced choice (2AFC) display where they were asked to determine whether the matching string was in the first or second display. The findings indicate that inconsistency is primarily responsible for the decrements in observed performance that can be seen over a time delay. Theoretical and practical applications are discussed.

In today's fast-paced society, the need to multitask is inevitable. Storing information in short-term memory (STM) is just one of the items on the long list of tasks individuals must perform daily. For example, while driving, an individual must remember the directions she has been given and be able to recognize the correct street names on her route. Even with the use of a navigation system, a driver must still hold that information in her memory long enough to find the appropriate street. Drivers must also temporarily store information presented on road signs, which restaurants are located at the upcoming exit, and the location of the nearest rest area. Even Amber Alerts displayed along highways and interstates rely on an individual's STM to recognize license plate numbers and vehicle descriptions in order to locate missing children.

Driving is just one example of the many daily tasks that require cognitive multitasking. There are countless circumstances in which individuals must utilize STM to enhance their performance on any given task. Considering the potentially infinite circumstances in which an individual may be required to utilize STM, along with the limited capacity of STM, it is imperative to investigate all factors which may be related to performance on STM tasks.

Short-Term Memory

The immediate memory store has been known by many different names including immediate memory, primary memory, working memory, and STM. While there are differences between these terms, those differences are not a main focal point of this paper. Thus, we will continue to use STM throughout to refer to the memory system that is holding information in conscious awareness. The concept of a limited capacity memory store was introduced by William James in 1890. James recognized a "primary memory store" that holds the information of the conscious mind. The mechanisms by which information is retrieved and lost from this store have engaged psychologists for over a century. Past research has shown that recall for information presented is strongest immediately after presentation and dissipates after a short interval of time (Daniels, 1985; Mutter, 1980; Peterson & Peterson, 1959).

Multiple models have been proposed to explain how STM operates (Atkinson & Shiffrin, 1968; Baddeley, 1986; Baddeley & Hitch, 1974; Broadbent, 1958; Cowan, 1988). In the basic STM paradigm, researchers ask participants to view a set of stimuli and then recall it after various time intervals. A number of studies have repeatedly shown that participants perform better on a memory task after a short delay than after a long delay (Daniels, 1985; Mutter, 1980; Peterson & Peterson, 1959).

One possible cause of inferior recall after a longer delay is that information erodes as time passes (Baddeley, 1986; Barrouillet, Bernardin, & Camos, 2004; Page & Norris, 1998; Portrat, Barrouillet, & Camos, 2008); a process Thorndike (1914) termed decay. Researchers have found that when participants were given two concurrent tasks, a memory task and a secondary task to prevent rehearsal, performance seemed to be a function of the length of the retention interval. Longer delays led to lower recall rates (Peterson & Peterson, 1959; Reitman, 1971; Shiffrin, 1973).

Another possible explanation of inferior recall after a longer delay is based on interference. Interference is when current information is lost from memory due to the introduction of new information. Interference can be proactive or retroactive. Proactive interference occurs when a person cannot recall information because previously learned information disrupts the retrieval process (Kepple & Underwood, 1962). Retroactive interference occurs when a person cannot recall information because new information disrupts the retrieval process (Barnes & Underwood, 1959; Wohldmann et al, 2008). In many cases, the secondary task intended to prevent rehearsal in the decay paradigm introduces new information that also requires STM resources. Thus, the decrement in recall could be due to the secondary task interfering with the information to be recalled (Reitman, 1974; Roediger, Knight, & Kantowitz, 1977).

The goal of this paper is not to test these theories against each other, but to introduce a third possible alternative explanation of inferior recall after a longer delay. Recent research in a variety of task performance paradigms has uncovered a factor that has been notably underemphasized in many areas of psychology and completely overlooked in many others. Consistency has shown to be a fundamental factor in task performance, time and time again (e.g. Hunt, Rice, Trafimow, & Sandry, 2013; Rice et al., 2012; Rice, Geels, Trafimow, & Hackett, 2011; Rice, Trafimow, & Hunt, 2010; Rice, Trafimow, Keller, Hunt, & Geels, 2011; Trafimow, Hunt, Rice, & Geels, 2011; Trafimow & Rice, 2008, 2009, 2011). We propose that a lack of consistency may be a potential cause of the commonly found decrement in performance after a long delay as compared to a short delay. In order to test this explanation, we will utilize a theory that is able to separate the effects of random factors from the effects of systematic factors and provide a precise measure of consistency. Potential Performance Theory (Trafimow & Rice, 2008, 2009, 2011) provides the methodology and mathematical formulas to test this new explanation.

Potential Performance Theory

Potential Performance Theory (PPT; Trafimow & Rice, 2008, 2009, 2011) is a general theory of task performance which mathematically parses random factors from systematic factors. From this, researchers are able to calculate each participant's consistency on a given task. The term consistency, as it relates to PPT, has a very specific definition: the correlation between participant responses on two blocks of identical trials. It is important to note that consistency is an inverse measure of randomness. In other words, as random noise increases, consistency will decrease and vice versa.

In order to conduct an analysis using PPT mathematics, participants must be presented with one block of trials and each trial must contain a stimulus that is different from any other trial in some way. Participants should then be given the same block of trials a second time, with each trial in the second block matching an identical trial from the first block. In other words, participants should be presented with each unique stimulus two times. This allows researchers to calculate a correlation between the responses on the first block of trials and the responses on the second block of trials. This correlation is termed the consistency coefficient. To reiterate, a low consistency coefficient indicates the presence of more randomness, whereas a high consistency coefficient indicates the presence of less randomness.

Using observed scores and the correlation between two blocks of trials, researchers can also calculate each participant's potential score. The potential score is the best estimate of the score each participant could have obtained in the absence of any randomness (i.e. given perfect consistency). (For a detailed explanation of the mathematical techniques and formulas used to perform a PPT analysis, see Trafimow & Rice, 2008, 2009, 2011.)

Current Study

As noted above, previous research on STM has primarily focused on systematic causes of decreased performance with a longer delay before recall. While many researchers continue the debate over decay and interference theories, our goal is not to provide support for any one theory. Rather, we aim to use PPT to reveal whether there is another important variable that may be contributing to the diminished proficiency of STM after a delay interval. We propose one basic hypothesis, followed by three possible explanations. First and foremost, we believe that observations from previous studies will be replicated, showing a decrement in performance with a longer (16 second) delay between observation and recall, as opposed to a shorter (2 second) delay. Given this hypothesis, we offer possible explanations for this effect.

The first explanation indicates that a systematic cognitive change may be occurring between the short and long delay conditions, causing the decrement in performance with the added delay. If this systematic change is the sole cause of the decrement in performance, one could expect to see a notable drop in potential scores, with consistency scores remaining relatively constant. It is important to note that in this case, the actual consistency level is less important than the predicted lack of change in consistency between conditions.

The second explanation indicates that an increase in the length of delay between observation and recall causes an increase in random noise, rather than a systematic variation. If an increase in random noise is the primary or sole cause of the decrement in performance between the two conditions, one could expect to observe a notable drop in consistency scores, with potential scores remaining relatively constant. It is again important to note that the actual potential score is less important than the predicted lack of change in potential scores between conditions.

Finally, the third possible explanation is that both random and systematic factors operate together to cause the expected drop in observed performance. In this situation, one would expect to see a notable change in both potential scores and consistency scores between the short and the long delay conditions.

METHOD

Participants

A total of 30 undergraduates (14 females, 16 males) from a large Southwestern university participated in this experiment to partially fulfill a course requirement. The mean age was 20.4 years (SD = 6.89 years, range = 18-56 years). Each participant had normal or corrected-to-normal visual acuity.

Apparatus and Stimuli

Stimuli for this study were presented on a Dell computer with a 22 in. (53.3 cm) LCD monitor set at 1024 x 768 resolution using E-Prime 2.0. Each stimulus consisted of a string of 7 characters randomly selected from the English alphabet. These characters were presented in Courier New 48 point black font in a straight horizontal line in the center of the screen. Each character took up approximately 1.75 (width) x 1.5 (height) degrees of visual angle.

Procedure and Design

Before beginning the experiment, participants were asked to sign consent forms and were given a test to ensure 20/20 vision. They were then comfortably seated directly in front of the experimental display and asked to place their chin in a chinrest for the duration of the experiment. The chinrest was used to control viewing distance from the screen at exactly 21 inches. Once comfortably situated, participants were presented with the instructions for the task on the computer screen and were given the opportunity to ask questions before starting the experiment.

Each trial began with a fixation cross in the center of the screen, displayed for 500 ms. Following the fixation, a string composed of seven randomly generated letters was displayed for 1,000 ms, immediately followed by a second fixation. One block of trials presented the fixation for 2 seconds, while the other block of trials presented the fixation for 16 seconds. In order to calculate a correlation coefficient used in the PPT analysis, each trial must appear two times, thus all participants saw two sets of each block for a total of four blocks of trials. All four blocks were presented in random order. After the fixation, two interval choices were displayed, beginning with the first interval choice presented for 1,000 ms, a third fixation for 500 ms, and then the second interval choice for 1,000 ms. Finally, participants were presented with the response screen. They were asked to press the "F" key to indicate that the matching string of letters was in the first interval or to press the "J" key to indicate that the matching string of letters was in the second interval. The response screen remained visible until a response was made.

Incorrect answers were generated by replacing one letter from the original stimulus with an alternate randomly selected letter. The position of the changed letter was also selected at random. Half of the trials in each block contained the correct answer in the first interval, while the other half of the trials in each block contained the correct answer in the second interval.

Each of the 4 blocks consisted of 50 trials, all presented in random order. All blocks were identical, with the exception of the length of delay between the original stimulus and the interval choices. Importantly, each trial within a block was unique and matched by only one trial in the other block of that condition, thereby allowing the computation of a consistency coefficient across the 50 matched trials in a pair of blocks. A within-participants design was used, in which all participants were presented with all trials across all four blocks.

RESULTS

Group Results

To analyze the data, we performed three paired samples *t* tests on observed scores, potential scores, and consistency, respectively. There was a significant difference in observed scores, $t(29) = 3.14$, $p = .004$, $d = 1.17$; not surprisingly, participants' observed performance was better in the short delay condition ($M = 0.85$, $SD = 0.09$) than in the long delay condition ($M = 0.80$, $SD = 0.10$). However, and interestingly, no significant difference was detected in the potential scores, $t(29) = 0.15$, $p = .88$, $d = 0.04$, demonstrating that in the absence of any random factors, participants could have potentially reached a ceiling effect in performance in both the short delay and the long delay conditions ($M = 0.99$, $SD = 0.06$; $M = 0.98$, $SD = 0.26$, respectively). Contrary to potential scores, consistency between conditions did reach significance, $t(29) = 2.53$, $p = .02$, $d = 0.94$, with a strong effect size. Participants were much more consistent when performing in the short delay condition ($M = 0.55$, $SD = 0.25$) than when performing in the long delay condition ($M = 0.47$, $SD = 0.19$).

Individual Results

One of the benefits of conducting a PPT analysis is that researchers are able to analyze individual participant data to determine the observed score, potential score, and consistency of each person's performance independently, rather than relying solely on group means. This gives researchers the ability to identify patterns that are specific to individuals and may not be clear from group data. For example, in this study, although the group data indicates that consistency is a major contributor to the decrement in performance, it is not clear just how much it can influence observed scores. Here, we will present individual analyses from three participants to demonstrate, in particular, the impact of consistency on observed performance.

Unlike the general trend of the group, Participant 8 showed no change in observed score between the short delay condition and the long delay condition ($M = .78$). While it may be tempting to conclude that there is in fact no change between conditions for this participant, that assumption would be misguided. In actuality, there are two separate effects in this case that each contrast the other and mask the changes that are taking place. First, despite the lack of change in the observed score, the potential score of Participant 8 actually increased between the short delay condition ($M = .95$) and the long delay condition ($M = 1.00$). His observed score remained unchanged, however, because of a simultaneous decrease in his consistency between the short delay condition ($M = .39$) and the long delay condition ($M = .29$). In essence, the decline in consistency was powerful enough to counteract the improvement of this participant with respect to the favorability of systematic factors.

Participants 19 and 27 are also prime examples to demonstrate just how influential consistency can be on observed scores. Each of these participants exhibited a decline in observed scores between the short delay condition ($M_{19} = .90$; $M_{27} = .88$) and the long delay condition ($M_{19} = .86$; $M_{27} = .80$), consistent with the group trend. Interestingly, though, both participants actually showed an improvement in potential scores between the short delay ($M_{19} = .94$; $M_{27} = .95$) and long delay (M19 = .98; M27 = 1.00) conditions. These improvements were not only masked, but actually counteracted by their drops in consistency.

DISCUSSION

In daily life, individuals are continuously tasked with a long list of cognitive demands that often must be performed simultaneously. One of these demands is to store a certain amount of information in STM, however, many researchers have shown that as more time elapses, recall for that information degrades. Researchers have proposed both decay and interference as potential explanations of this degradation. The purpose of this paper was to introduce a third possible explanation for this effect: inconsistency.

Derived from previous research, we hypothesized that we would replicate the finding that performance diminishes when participants are presented with a longer delay between observation and recall. Not surprisingly, we did replicate these findings, showing a marked decrease in performance on the long delay condition as compared to the short delay condition. Based on PPT, we proposed three possible explanations for these findings. First, we suggested that a systematic cognitive change may be the primary or sole cause of the decrement in performance between the two conditions, which would result in a notable drop in potential scores, with consistency remaining relatively constant. Alternatively, we suggested that an increase in the length of delay causes an increase in random noise instead of a systematic variation, which would result in a notable drop in consistency, with potential scores remaining relatively constant. Finally, the third possible explanation was that both systematic cognitive changes between the short and long delay conditions as well as an increase in random noise operate together to cause the expected drop in observed performance, resulting in a notable change in both potential scores and consistency between the short and the long delay conditions.

Results for this study were clear and straightforward. After mathematically parsing participants' observed performance into two underlying components—a potential score and a consistency score—we found that potential scores did not differ from each other as a function of the delay period; potential scores in the short delay condition mirrored the potential scores in the long delay condition. In contrast, we found a significant difference in consistency that was solely responsible for driving the diminished performance between the two delay conditions, supporting the second explanation of the findings. In other words, had participants been perfectly consistent, their performance would have held steady at nearly perfect accuracy across all blocks, regardless of whether they were presented with a short or a long delay.

This information may well prove to be extremely valuable, not only theoretically, but in applied settings as well. Many professions rely heavily on individuals' ability to store information in STM and recognize that information when it is presented at a later point in time. For example, when an Amber Alert is issued for a missing child, the license plate of the accused is often available. Police officers are at a distinct disadvantage if they are unable to recognize the license plate should they come across it. The ability to recognize the string of numbers and letters without constantly rechecking the Amber Alert information would undoubtedly be a major benefit for police officers as well as the missing child.

The data provided in this study can easily be used as a guide to determine the type of training that would be most effective in increasing recognition abilities. The observed data in this study suggest that individuals are not sufficiently skilled in this area. Because potential scores in this study were near ceiling, we can infer that individuals are in actuality quite good at this task. It is their inconsistency that is masking this high level of performance. A PPT analysis suggests that individuals need not be trained to improve their skill in recognition tasks, but instead they should be trained to learn to use their (already quite impeccable) skill more consistently.

In addition, in looking at individual performances, we were able to investigate how much inconsistency can drive down observed performance. To show the tremendous power consistency has over observed scores, we were able to identify cases in which, with perfect consistency, participants would have actually shown an opposite effect from the group data. This did not translate to the observed data, however, because the drop in consistency in each case was sufficient to either mask or reverse the systematic improvement. The ability to analyze data individually, rather than being restricted to group averages, allows the additional advantage of identifying where each person is lacking. This means that training techniques can be uniquely tailored to the needs of each individual. Those who are quite consistent, but systematically inaccurate may benefit from a different training plan than those who are quite systematically accurate but lacking consistency, or those who are both systematically inaccurate and also lacking consistency.

One may wonder why consistency should be considered an alternative to decay and interference, as opposed to simply a different way to measure what has already been presented in the literature for decades. We hope to make this distinction quite clear. Decay theory posits that performance on an STM task degrades with a longer delay because the information tends to erode over time. Interference theory suggests that performance on an STM task degrades with a longer delay because other information interferes with the ability to actively rehearse the target information stored in STM. In the case of decay theory, the only way to correct this problem is to reduce the amount of time elapsed between presentation and recall. In the case of interference theory, the only way to correct this problem is to reduce the amount of other information stored in STM. In either situation, drastic and unrealistic steps must be taken to improve performance. By using PPT, on the other hand, we have uncovered a factor that shows the potential for individuals to greatly increase their performance on an STM task simply by learning to use their recall strategies more consistently.

While utilizing the mathematical techniques provided by PPT allowed us to investigate an additional consideration for STM research, it is also necessary to address its limitations. PPT provides the ability to determine the extent to which random and systematic factors each play a role in observed performances, and, more specifically, in the differences in observed performances across conditions. PPT does not, however, allow the researcher to identify the specific factors that are at play. For example, while we can see in this study that the change in observed performance is due to an increase in random noise, had there also been a decrease in potential scores, we would not have been able to identify the specific cognitive mechanism contributing to this change. In the future, researchers who are investigating the difference between alternative STM theories may wish to utilize PPT methodology in conjunction with strong theorizing to explore these issues.

Although PPT does have its limitations, it is important that its contributions are not minimized. The present study has introduced a new variable, consistency, that has otherwise not been considered and one that this study has shown to be a contributing factor to the decrement in STM performance over a delay. This study has tremendous potential implications for future research on cognitive tasks. First, it gives researchers a new path to follow in an attempt to improve individuals' STM performance; a path that is neither impossible nor unrealistic. Second, it provides a new variable that may be explored in conjunction with known systematic effects that have the potential to make a drastic impact on existing theories. For example, using PPT methodology to consider the effects of consistency alongside a previously measured systematic variable would allow researchers to identify how much of the observed effect is truly due to that variable and how much of the effect can be accounted for by a lack of consistency. Finally, given that PPT is a general theory of task performance, this new finding opens the door for exploration of cognitive tasks well beyond the STM paradigm. PPT can potentially be used to investigate any of the numerous tasks that vehicle operators must perform on a regular basis.

REFERENCES

Allen, M. J., & Yen, W. M. (1979). Introduction to measurement theory. Monterey, CA: Brooks/Cole Publishing Company.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), The Psychology of Learning and Motivation: Advances in Research and Theory. New York: Academic Press.

Baddeley, A. D. (1986). Working Memory. Oxford, England: Clarendon Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? Trends in Cognitive Science, 4, 417–423.

Baddeley, A. D. (2002). Is working memory still working? European Psychologist, 7, 85–87.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), The psychology of learning and motivation: Advances in research and theory. New York: Academic Press.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. Journal of Verbal Learning and Verbal Behavior, 14, 575–589.

Barnes, J. M., & Underwood, B. J. (1959). Fate of first list association in transfer theory. Journal of Experimental Psychology, 58, 97–105.

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. Journal of Experimental Psychology, 133, 83–100.

Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 33, 570–585.

Berman, M. G., Jonides, J., & Lewis, R. L. (2009). In search of decay in verbal short-term memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 35, 317–333.

Broadbent, D. E. (1958). Perception & Communication. London: Pergamon Press.

Cohen, R. J., & Swerdlik, M. E. (1999). Psychological testing and assessment: An introduction to tests and measurements (4th ed.). Mountain View, CA: Mayfield.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. Psychological Bulletin, 104, 163–191.

Crocker, L., & Algina, J. (1986). Introductions to Classical and Modern Test Theory. New York: Holt, Rinehart and Winston.

Daniels, A. H. (1895). The memory after-image and attention. American Journal of Psychology, 6, 558–564.

Hunt, G., Rice, S., Trafimow, D., & Sandry, J. (2013). Using potential performance theory to analyze systematic and random factors in enumeration tasks. American Journal of Psychology, 26(1), 23-32.

James, W. (1890). The Principles of Psychology. NY: Henry Holt.

Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. Journal of Verbal Learning and Verbal Behavior, 1, 153–161.

Lewandowsky, S., Duncan, M., & Brown, G. D. A. (2004). Time does not cause forgetting in short-term serial recall. Psychonomic Bulletin & Review, 11, 771–790.

Lord, F. M., & Novick, M. R. (1968). Statistical theories of mental test scores. Reading, Massachusetts: Addison-Wesley Publishing Company.

McKone, E. (1998). The decay of short-term implicit memory: Unpacking lag. Memory & Cognition, 26, 1173–1186.

Mueller, S. T., Seymour, T. L., Kieras, D. E., & Meyer, D. E. (2003). Theoretical implications of articulatory duration, phonological similarity, and phonological complexity in verbal working memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 29, 1353–1380.

Muter, P. (1980). Very rapid forgetting. Memory & Cognition, 8, 174–179.

Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. Annual Review of Psychology, 53, 53–81.

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. Psychological Review, 105, 761–781.

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. Journal of Experimental Psychology, 58, 193–198.

Portrat, S., Barrouillet, P., & Camos, V. (2008). Time-related delay or interference-based forgetting in working memory? Journal of Experimental Psychology: Learning, Memory, and Cognition, 34, 1561–1564.

Postle, B. R., D'Esposito, M., & Corkin, S. (2005). Effects of verbal and nonverbal interference on spatial and object visual working memory. Memory & Cognition, 33, 203–212.

Reitman, J. S. (1971). Mechanisms of forgetting in short-term memory. Cognitive Psychology, 2, 185–195.

Reitman, J. S. (1974). Without surreptitious rehearsal, information in short-term memory decays. Journal of Verbal Learning and Verbal Behavior, 13, 365–377.

Rice, S., Geels, K., Hackett, H., Trafimow, D., McCarley, J., Schwark, J., & Hunt, G. (2012). The harder the task, the more inconsistent the performance: A PPT analysis on task difficulty. Journal of General Psychology, 139, 1-18.

Rice, S., Geels, K., Trafimow, D., & Hackett, H. (2011). Our students suffer from both lack of knowledge and consistency: A PPT analysis of test-taking. US-China Education Review.

Rice, S., Trafimow, D., & Hunt, G. (2010). Using PPT to analyze sub-optimal human-automation performance. Journal of General Psychology, 137(03), 310–329.

Rice, S., Trafimow, D., Keller, D., Hunt, G., & Geels, K. (2011). Using PPT to correct for inconsistency in a speeded task. The Journal of General Psychology, 138(1), 12–34.

Roediger, H. L., Knight, J. L., & Kantowitz, B. H. (1977). Inferring decay in short-term memory: Issue of capacity. Memory & Cognition, 5, 167–176.

Schweickert, R., & Boruff, B. (1986). Short-term memory capacity: Magic number or magic spell? Journal of Experimental Psychology: Learning, Memory, and Cognition, 12, 419–425.

Shiffrin, R. M. (1973). Information persistence in short-term memory. Journal of Experimental Psychology, 100, 39–49.

Thorndike, E. L. (1914). Educational Psychology: Briefer Course. New York: Teachers College, Columbia University.

Trafimow, D., Hunt, G., Rice, S., & Geels, K. (2011). Using potential performance theory to test five hypotheses about meta-attribution. Journal of General Psychology, 138(2), 1–13.

Trafimow, D., & Rice, S. (2008). Potential Performance Theory: A general theory of task performance applied to morality. Psychological Review, 115(2), 447–462.

Trafimow, D., & Rice, S. (2009). Potential Performance Theory (PPT): Describing a methodology for analyzing task performance. Behavior Research Methods, 41(2), 359–371.

Trafimow, D., & Rice, S. (2011). Using a sharp instrument to parse apart strategy and consistency: An evaluation of PPT and its assumptions. Journal of General Psychology, 138(3), 169-184.

Wixted, J. T. (2005). A theory about why we forget what we once knew. Current Directions in Psychological Science, 14, 6–9.

Wohldmann, E. L., Healy, A. F., Bourne Jr., L. E. (2008). A mental practice superiority effect: Less retroactive interference and more transfer than physical practice. Journal of Experimental Psychology: Learning, Memory, and Cognition, 34, 823–833.