© Journal of the Indian Academy of Applied Psychology
January 2010, Vol.36, No.1, 151-158.

Performance Feedback, Mental Workload and Monitoring Efficiency

Anju L. Singh, Trayambak Tiwari, and Indramani L. Singh
Banaras Hindu University, Varanasi

The present study examined the effect of success and failure performance feedback on perceived mental workload and monitoring performance in flight simulation task. The revised version of the multi-attribute task battery (MATB) was administered on 20 non-pilot participants. The performances were recorded as hit rates, false alarms and root mean square errors. Mental workload was assessed using NASA-TLX questionnaire. A 2(success-failure feedback) x 2(30-min sessions) x 3(10-min blocks) mixed factorial design with repeated measures on last two factors was used. The obtained results revealed that performance feedback did not have a significant effect on mental workload and malfunction detection. The findings support the notion that monitoring inefficiency (i.e., automation-induced complacency) is a robust phenomenon and it can be observed in multi-task environment with high static automation reliability.

Keywords: Mental workload, Feedback, Monitoring performance, Automation-induced complacency, Flight simulation task

The technological revolution has gradually removed the human-operators of many complex systems from front-line levels of control and having their actions relayed via an intervening mass of computers and microprocessors. Instead of active controller of the system, the operator of an automated system has now become a passive observer. It may seem paradoxical, but automated systems can both reduce and increase mental workload. Therefore, mental workload is considered an important factor in the area of automation research. One of the fundamental reasons for introducing automation in complex systems is to reduce workload, and thereby to reduce human error. However, evidence shows that this is not necessarily true in all situations. Infact the automation merely changes how work is accomplished (Woods, 1994). Further, Reinartz and Gruppe (1993) argued that automated system present cognitive demands, which increases workload. The performance of the operator is hindered by the increase in processing load resulting from the additional task of collecting information about the system state. This is further complicated by the extent of the operator's knowledge about the system. In the event of manual takeover, the operator must either disable interlocks to other systems, or else match his/her actions to those of related process functions.

Operators can use different strategies to cope with workload. Hart (1989) showed that experienced operators work in advance during periods of low workload in order to eliminate workload peaks in the future. Hockey (1993) has presented different strategies to cope with workload in a regulation model. Operators constantly compare their performance with the goal state. If the quality of the performance is not good enough according to the goal state, more effort will be invested. To a certain level this is an automatic process. An effort monitor evaluates the amount of effort that is required and when the effort increases too much, the

performance evaluation process is controlled at higher cognitive level. Operators can apply different strategies to situations in which the performance level does not fit the goal state. They can decide to invest more effort or to decrease the goal and accept a lower level of performance. When these strategies are not possible because the performance level is already low or the operator has already invested a maximum amount of effort, the situation leads to stress. Gaillard and Wientjes (1994) have shown that there are substantial costs involved when one has to invest a lot of mental effort to perform a highly demanding task.

It is well understood that feedback or knowledge of result (KR) is a crucial factor in the early stages of skill acquisition (e.g., Groeger, 1997). This has been applied to many diverse fields from consumer products (Bonner, 1998) to aviation (White, Selcon, Evans, Parker, & Newman, 1997). In the latter study, it was found that providing redundant information from an additional source can actually elicit a performance advantage.

One study that is relevant to the driving domain examined the effects of feedback on performance of controlled and automatic tasks was conducted by Tucker and associates (Tucker, MacDonald, Sytnik, Owens, & Folkard, 1997) and it was found that feedback can reduce error rates on tasks requiring controlled processing. However, automatic tasks were found to be resistant to the effects of feedback. Furthermore, a vigilance decrement was observed only in the controlled task, suggesting automatic responses do not suffer from such a decrement. This vigilance decrement was also found to be unaffected by feedback.

Researches have also been conducted to compare novice and expert drivers in the context of automation. For example, Duncan, Williams and Brown (1991) examined the performance of a group of normal (experienced) drivers with that of novices and experts on a subset of driving skills. They found that on half of the measured skills, the normal drivers actually performed worst than novice drivers, who performed at a similar level to the experts. Results revealed that the normal (experienced) drivers succumbed to a range of bad habits in the absence of learning feedback.

In complex systems, such as modern fighter jets and helicopters, operators have to manage several tasks at the same time that increased pilots' mental workload. An important and challenging problem in many multi-task environments is managing interruption (McFarlane & Latorella, 2002). Researchers have noted that proactive systems executing in environments such as aviation cockpits (Dismukes, Young, & Sumwalt, 1998; Latorella, 1996), control rooms (Stanton, 1994), in-vehicle displays (Lee, Hoffman, & Hayes, 2004) and office environments (Bailey & Konstan, 2006; Czerwinski, Cutrell, & Horvitz, 2000b; Jackson, Dawson, & Wilson, 2001) are significantly interrupts the user's primary tasks'. It has also been observed that when primary tasks are interrupted at random moments, users take longer to complete the tasks (Bailey & Konstan, 2006; Czerwinski, Cutrell, & Horvitz, 2000a; Rubinstein, Meyer, & Meyer, 2001), commit more errors (Kreifeldt & McCarthy, 1981; Latorella, 1996) and experience increased levels of frustration, annoyance and anxiety (Adamczyk & Bailey, 2004; Bailey & Konstan, 2006; Zijlstra, Roe, Leonora, & Krediet, 1999).

The foregoing review, though, suggest that performance feedback is related with mental workload, but it is still a controversial issue that how and to what extant the performance feedback influences the mental workload. For example, Becker, Warm, Dember and Hancock (1995) found that performance feedback generally lowered mental workload in a monitoring task, whereas, the results of Fairclough, May and

Carter (1997) suggest that time headway feedback had no effect on workload in a car-following scenario. In the light of this inconsistency in findings the present study makes an attempt to examine how and to what extent the performance feedback is related with mental workload in multi-task situation.

It was hypothesized that the success feedback would reduce monitoring inefficiency more than failure feedback and participants would perceive low mental workload in success feedback condition than in failure feedback, resulting in low monitoring efficiency.

Method

Participants:

Participants in this study were 20 students of the Banaras Hindu University. Each participant had normal (20/20) or corrected to normal visual acuity, and their age varied from 18 to 23 years. None of the participants had prior experience on the flight simulation task.

Experimental Design:

A 2(feedback) x 2(session) x 3(block) mixed factorial design was employed in this experiment. Between-subjects variable had two levels of feedback i.e., success feedback and failure feedback, whereas within-subjects variables included sessions and blocks. Participants were randomly assigned in each experimental condition (success and failure feedback; n = 10 in each).

Tools:

Mental Workload Questionnaire: Participants completed the NASA-TLX (Hart & Staveland, 1988) before beginning the experiment. The NASA-TLX has six components reflecting the degree of mental demand, physical demand, temporal demand, performance, effort and frustration associated with a task. This scale provides an overall workload score based on a weighted average of ratings

Flight Simulation Task: A revised version of multi-attribute task battery (MATB: Comstock & Arnegard, 1992) with high automation reliability (87.5%) was used in this study. Automation reliability was defined as the percentage of correct detection of malfunctions by the automation routine in each 10-min block in the system-monitoring task. This is a multi-task flight simulation package comprising system-engine monitoring, compensatory tracking, fuel resource management, communications, and scheduling tasks. In the present study, only the system-engine monitoring, tracking, and fuel-resource management tasks were used, in which system-monitoring task was automated during test sessions. These three tasks were displayed in separate windows on a 14" colour monitor (For details regarding the task see: Singh, Sharma & Singh, 2005; Singh & Singh, 2006).

Procedure:

Upon arrival at the lab, participants were required to fill out a consent form and background questionnaire. The Snellen Eye Chart was used to test visual acuity of the participants. This test measures how well participants see at various distances. The participants were then asked to complete the pre-task NASA-Task Load Index. After completing the questionnaires, the experimenter provided a brief introduction about the flight simulation task to participants. In both of the experimental conditions, participants first completed a 10 -minutes practice, which allowed them to become accustomed to the task before participating in final test session. The correct and incorrect detection were recorded as the dependent measures for the system monitoring task and the root mean square errors were recorded for the tracking and the fuel management tasks. Participants, who score above 60% on hit rates, were eligible for a final six 10-minutes test blocks. In test session, system-monitoring task was automated, and the

participant has to perform manually tracking and resource management tasks. However, automation will not be 100% reliable, so they have to keep their eyes on system-monitoring task and, if automation fails then they have to fix it by pressing designated keys from the keyboard. After the termination of the task, participants completed the post-task NASA-Task Load Index, with specific reference to the flight simulation task. Feedback on performance for all three tasks was given to each participant as per the design. The entire experiment was completed in approximately 1 hour and 30 min.

Results and Discussion

Practice Performance

Results of practice session indicated that the mean correct detection (hits) of all subjects varied from 60% to 80%, irrespective of feedback conditions. However, hit rates performance of participants didn't significantly differ from success to failure feedback condition. Similar results were also obtained for remaining dependent measures like false alarms, tracking and fuel resource management. The finding demonstrates that all the participants have comparable level of performance on the experimental task before entering into the final experiment.

Correct Detection Performance (hit rates)

Means and SDs for correct detection of malfunctions on system-monitoring task were computed for each of the two sessions i.e., before and after manipulation of feedback performance. Mean performances indicated that participants obtained high detection of malfunction scores (M = 80.47; SD = 22.78) in the first session than in the second session (M = 67.13; SD = 17.81) under success feedback condition. Similarly, participants achieved high mean scores in the first session (M = 83.53; SD = 17.16) than in the second session (M = 48.37; SD = 26.74) under failure feedback condition. The total mean performance across six 10-min blocks for the success feedback was higher (M = 73.80; SD = 20.29) than in the failure feedback (M = 65.95; SD = 21.95). Results further revealed that participant's monitoring efficiency reduced across blocks, irrespective of the feedback types.

Correct monitoring performance data were then submitted to a 2(feedback) x 2(session) x 3(block) analysis of variance with repeated measures on the last two factors for examining interaction effect, if any. The ANOVA results showed that the main effect of feedback was not found significant, which revealed that the types of feedback either success or failure given prior to the detection of automation failures had no impact on monitoring performance. So these results do not support our first hypothesis that the success or failure feedback performance would reduce monitoring performance like on other psycho-motor task performances. Moreover, the main effect of session was found significant ($F_{(1, 18)} = 23.04$; $p<0.01$), which revealed that participants monitoring performance was significantly deteriorated across sessions. Similarly, the interaction between feedback and session was also found significant ($F_{(1, 18)}=4.67$; $p<.05$) (see Figure-1).

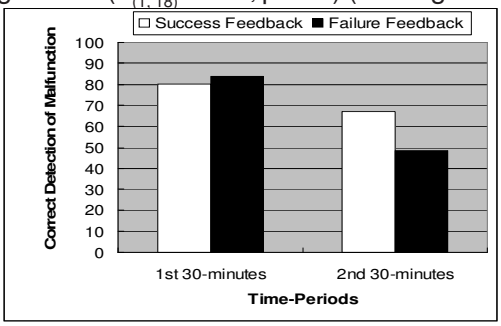


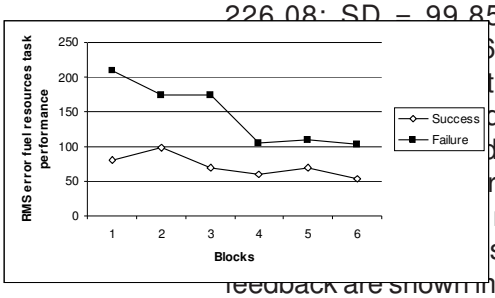
Figure 1: Correct detection of malfunction (Hit rates performance) as function of feedback and session

This interaction effect could be due to steep decrement in hit rates over sessions. Thus, the obtained results deciphered that detection of automation failures (monitoring

performance) would progressively decline over sessions. This finding is consistent with other researchers (Parasuraman, Molloy, & Singh, 1993; Singh, Molloy, & Parasuraman, 1997), who reported monitoring inefficiency over time periods under multi-task environment, especially in monitoring automation failures.

Tracking Performance

Means and SDs for integrated RMS error on the tracking task were calculated and results indicated that participants obtained high tracking performance scores (M = 234.74; SD = 98.37) in the first session than in the second session (M = 217.42; SD = 101.31) under success feedback condition. Similarly, participants achieved high mean scores in the first session (M = 227.61; SD = 81.49) than in the second session (M = 203.70; SD = 83.98) under failure feedback condition. The total mean performance across six 10-min blocks for the success feedback was higher (M = 226.08; SD = 99.85) than in the failure [illegible]; SD = 82.73). Results [illegible] participant's tracking [illegible] across blocks in the [illegible] condition while it remained [illegible] periods success feedback [illegible] mean square error [illegible] success and failure feedback are shown in Figure- 2. The ANOVA for the tracking performance showed an improvement over sessions ($F_{(1, 18)} = 4.59$; $p < .05$). However, feedback of any type did not affect tracking performance as a whole.



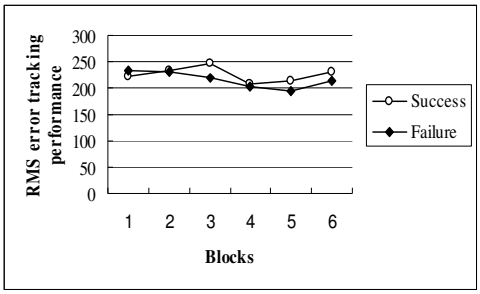


Figure 2: Tracking performance as function of feedback and block

Fuel resource management performance

Mean performance on the fuel management task across six 10-min blocks indicated that participants obtained high scores in the first session (M = 82.48; SD = 76.92) than in the second session (M = 60.20; SD = 30.95) in success feedback condition. Similarly, participants achieved high mean scores in the first session (M = 186.42; SD = 190.61) than in the second session (M = 105.85; SD = 99.43) under failure feedback condition. The total mean performance across six 10-min blocks for the failure feedback was higher (M = 146.14; SD = 145.02) than in the success feedback (M = 71.34; SD = 53.94) (see Figure 3). Analysis of variance for the fuel management performance showed an improvement over sessions, like tracking performance ($F_{(1, 18)} = 4.83$; $p < .05$). However, feedback performance didn't influence fuel resource management performance across time periods. Thus, the obtained results indicate that the types of feedback do not affect on tracking and on fuel management performance, especially in multiple tasks environment.

Figure 3: Fuel resource performance as function of feedback and block

Feedback and Workload

To examine the second hypothesis that success feedback would reduce workload, causing monitoring inefficiency. Means and standard deviations for each component of mental workload for success and failure feedback were computed and are presented in the Table-1 and these scores are also graphically displayed in the Figure- 4.

Mean workload scores indicated that pre-mental demand, pre-temporal demand, pre-effort and pre-frustration were higher than their counterparts of the post-mental workload in success feedback condition. Similarly, pre-workload mean scores on the mental demand and effort workload were also found higher in failure feedback condition than post-mental demand and post-effort workload. The mean differences between pre- and post-test sessions on the various components of the mental workload were further compared by using paired t-test. Statistically significant difference found only for effort workload subscale from pre- to post-session (t=2.23; p<0.05) at success feedback condition. The result revealed that participants in the success feedback condition rated the task as more effortful. Whereas, participants perceived significantly higher performance workload from pre- to post session (t=3.51; p<0.01) at failure feedback condition. None of the other subscales scores of mental workload were found to be statistically significant (p>0.05) at success and failure feedback conditions. These findings suggest that success and failure feedback have no impact on perceived mental workload across sessions. Thus, the obtained findings do not support our assumption that feedback would reduce mental workload, while performing multi-task in high static automation reliability condition, causing monitoring inefficiency.

Table 1: Mean, SD and paired sample t-values at pre- and post-test sessions for the various components of mental workload under success and failure feedback conditions

Conditions	Success Feedback(N=10)			Failure Feedback(N=10)		
	Mean	SD	t-value	Mean	SD	t-value
Pre-mental demandsvsPost-mental demands	79.00 78.60	10.21 13.40	.08	84.50 73.00	15.53 22.13	1.97
Pre-physical demandsvsPost-physical demands	36.70 28.28	21.14 59.50	-.91	48.60 20.60	22.74 41.50	1.54
Pre-temporal demandsvsPost-temporal demands	66.50 60.50	30.18 24.77	.77	54.50 57.50	28.13 24.63	-.41
Pre-effortvsPost-effort	72.50 61.60	24.74 26.47	2.33	82.00 65.00	15.31 31.39	1.96
Pre-frustrationvsPost-frustration	28.70 15.90	23.68 14.53	1.59	21.00 21.30	14.18 20.05	-.09
Pre-performancevsPost-performance	67.50 73.60	7.90 12.85	-1.80	63.40 78.60	10.90 10.14	-3.51

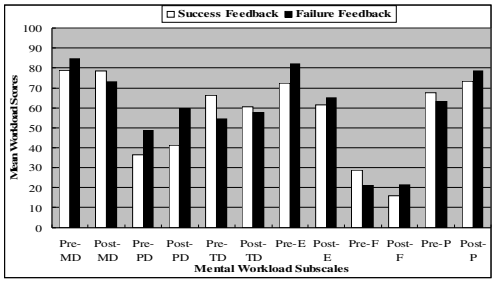


Figure 4: Mean workload scores as function of pre and post workload subscales and success and failure feedback conditions (MD=Mental Demand; PD=Physical Demand; TD=Temporal Demand; P=Performance; E=Effort; F=Frustration)

In sum, the findings suggest that the feedback performances would not facilitate monitoring efficiency and mental workload in multi-tasks environment. This result corroborates the findings of Singh, Hilburn and Parasuraman (1999), who also could not find any significant effect of online feedback on automation-induced complacency.

It appeared that the feedback manipulation had small and statistically insignificant benefits on system-engine monitoring, tracking, fuel resource management and mental workload

performance measures. However, this benefit could be more realized on manual task performance. This result is in line with findings of the Tucker et al. (1997) study.

We believe that the results of this study bring relevant information for the research on mental workload of the human operators in multi-tasks scenarios. A larger sample and alternative physiological indicators of mental workload could however be used in future researches aiming to study further the relations between feedback and monitoring performance in multi-task scenarios.

Future research is needed to explore further the precise relations between different types of feedback and their impact on human monitoring behaviour. Data of this nature could lay the basis for the development of a complete checklist about the human monitoring behaviour which would minimize the probability of an accident and enhances the efficiency of human operators.

References

Adamczyk, P. D., & Bailey, B. P. (2004). If not now when? The effects of interruptions at different moments within task execution. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 271-278.

Bailey, B. P., & Konstan, J. A. (2006). On the need for attention aware systems: Measuring the effects of interruption on task - performance, error rate, and affective state. *Computers in Human Behavior, 22*, 685-708.

Becker, A. B., Warm, J. S., Dember, W. N., & Hancock, P. A. (1995). Effects of jet engine noise and performance feedback on perceived workload in a monitoring task. *International Journal of Aviation Psychology*, *5*, 49-62.

Bonner, J. V. H. (1998). Towards consumer product interface design guidelines. In N. A. Stanton (ed.), *Human Factors in Consumer Products* (pp. 239- 258). London: Taylor & Francis.

Comstock, J. R., & Arnegard, R. J. (1992). *The multi-Attribute Task battery for human operator workload and strategic behaviour research* (Tech. Memorandum No. 104174). Hampton, VA: NASA Langley Research Center.

Czerwinski, M., Cutrell, E., & Horvitz, E. (2000a). Instant messaging and interruption: Influence of task type on performance. *Annual Conference of the Human Factors and Ergonomics Society of Australia (OZCHI)*, 356-361.

Czerwinski, M., Cutrell, E., & Horvitz, E. (2000b). Instant messaging: Effects of relevance and timing. *People and Computers XIV: Proceedings of HCI*, 71-76.

Dismukes, K., Young, G., & Sumwalt, R. (1998). Cockpit interruptions and distractions. *ASRS Directline, 10*.

Duncan, J., Williams, P., & Brown, I. (1991). Components of driving skill: Experience does not mean expertise. *Ergonomics*, *34*, 919-937.

Fairclough, S. H., May, A. J., & Carter, C. (1997). The effect of time headway feedback on following behaviour. *Accident Analysis and Prevention*, *29*, 387-397.

Gaillard, A.W. K. & Wientjes, C. J. E. (1994). Mental load and work stress as two types of energy mobilization. *Work & Stress, 8*, 141-152.

Groeger, J. A. (1997). *Memory and remembering: Everyday memory in context.* Harlow, UK: Longman.

Hart, S. G. (1989). Crew workload-management strategies: A critical factor in system performance.Proccedings of fifth International Symposium on Aviation Psychology, Columbus, OH, 22-27.

Hart, S. G., & Staveland, L. E. (1988). Development of the NASA-Task Load Index (NASA-TLX); Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (eds.), *Human mental workload* (pp.139-183). Amsterdam: Elsevier.

Hockey, G.R.J. (1993). Cognitive energetic control mechanisms in the management of work demands and psychological health. In A. D. Bradely & L. Weiskrantz (Eds.), Attention, selection awareness, and control: A tribute to Donald Broadbent (pp.328-345), Oxford University Press.

Jackson, T. W., Dawson, R. J., & Wilson, D. (2001). The cost of email interruption. *Journal of Systems and Information Technology, 5*, 81-92.

Kreifeldt, J. G., & McCarthy, M. E. (1981). Interruption as a test of the user-computer interface. *Proceedings of the 17th Annual Conference on Manual Control*, 655-667.

Latorella, K. A. (1996). Investigating interruptions: An example from the flight deck. *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting*, 249-253.

Lee, J. D., Hoffman, J. D., & Hayes, E. (2004). Collision warning design to mitigate driver distraction. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 65-72.

McFarlane, D. C., & Latorella, K. A. (2002). The scope and importance of human interruption in HCI Design. *Human-Computer Interaction, 17*, 1-61.

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

Reinartz, S. J., & Gruppe, T. R. (1993). Information requirements to support operator automatic co-operation. *Human Factors in Nuclear Safety Conference*, London, 22-23 April, 1994.

Rubinstein, J. S., Meyer, D. E., & Meyer, D. E. (2001). Executive Control of Cognitive Processes in Task Switching. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 763-797.

Singh, I. L. & Singh, A. L., (2006). Effects of auto training and static automation reliability on flight simulation monitoring task performance. *Indian Journal of Applied Psychology*, *43*, 35-40.

Singh, I. L., Sharma, H. O., & Singh, A. L. (2005). Effect of training on workload in flight simulation task performance. *Journal of Indian Academy of Applied Psychology, 31*, 81-90.

Singh, I. L., Hilburn, B., & Parasuraman, R. (1999). Effect of feedback on adaptive automation. *Journal of Indian Academy of Applied Psychology, 25*, 157-165.

Singh, I. L., Molloy, R. & Parasuraman, R. (1997). Automation-induced monitoring inefficiency: Role of display location. *International Journal of Human Computer Studies, 46*, 17-46.

Stanton, N. (Ed.). (1994). *Human Factors in Alarm Design*. London: Taylor and Francis.

Tucker, P., Macdonald, I., Sytnik, N. I., Owens, D. S., & Folkard, S. (1997). Levels of control in the extended performance of a monotonous task. In S. A. Robertson (ed.), *Contemporary Ergonomics* (pp. 357-362). London: Taylor & Francis.

White, J., Selcon, S. J., Evans, A., Parker, C., & Newman, J. (1997). An evaluation of feedback requirements and cursor designs for virtual controls. In D. Harris (ed.), *Engineering Psychology and Cognitive Ergonomics* (pp. 65-71). Aldershot, UK: Ashgate.

Woods, D. D. (1994). Automation: Apparent simplicity, real complexity. In M. Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: Current research and trends* (pp. 1-7). Hillsdale, NJ: Lawrence Erlbaum Associates.

Zijlstra, F. R. H., Roe, R. A., Leonora, A. B., & Krediet, I. (1999). Temporal factors in mental work: Effects of interrupted activities. *Journal of Occupational and Organizational Psychology, 72*, 163-185.

Received: September 27, 2009
Revision received: November 16, 2009
Accepted: December 16, 2009

Acknowledgement: This research was supported by a research grant from the Defense Research Development Organization (DRDO), New Delhi to Prof. Indramani L. Singh.

Anju L. Singh, PhD, Cognitive Science Laboratory, Department of Psychology, Banaras Hindu University, Varanasi-221 005**,** Email:anjubhu@rediffmail.com

Trayambak Tiwari, Research Scholar, Cognitive Science Laboratory, Department of Psychology, Banaras Hindu University, Varanasi-221 005.

Indramani L. Singh, PhD, Professor, Cognitive Science Laboratory, Department of Psychology, Banaras Hindu University, Varanasi-221 005.