Spontaneous and reflexive eye activity measures of mental workload

David L. Neumann & Ottmar V. Lipp

School of Psychology, The University of Queensland, Australia

Running head:
Eye activity and mental workload

Address for correspondence:
David L. Neumann, PhD, School of Psychology, The University of Queensland, QLD, 4072, Australia. E-mail: d.neumann@psy.uq.edu.au

Acknowledgements:
Grant A10027218 from the Australian Research Council supported this work.
Abstract

Measures of eye activity, such as blink rate and scanning patterns, have been used extensively as psychophysiological indices of mental workload. In a review of measures derived from spontaneous eye activity it is shown that different measures are differentially sensitive to specific aspects of mental workload. A less well known measure of non-spontaneous eye activity, the blink reflex, is also reviewed. Experiments using discrete punctuate stimuli and continuous tasks analogous to real-world systems show that blink reflexes are modulated by attention and that this modulation reflects modality-specific attentional engagement. Future research should examine the utility of the blink reflex according to the desirable properties of sensitivity, diagnosticity, validity, reliability, ease of use, unobtrusiveness, and operator acceptance.
The measurement of mental workload is an important consideration in human factors research because workload measures have the potential to assist in system design and development. Excessively high levels of mental workload can lead to overload, whereas underload can lead to complacency (Braby, Harris, & Muir, 1993) both of which may result in errors and can jeopardise the operator’s safety. Measures of eye activity have been extensively used in human factors research as an objective and non-invasive index of mental workload (e.g., Van Orden, Limbert, Makeig, & Jung, 2001; Veltman & Gaillard, 1998). In the context of mental workload assessment, eye activity is an encompassing term that includes various response components, such as endogenous eye blink, pupil diameter, and scanning patterns. Each of these measures has shown to be sensitive to workload demands in complex control tasks. A measure of eye activity that has only recently been explored in relation to workload assessment is the eye blink reflex. The eye blink reflex, in combination with other measures of eye activity, may prove to be a useful addition to the arsenal of mental workload measures. This is particularly relevant when one considers that the multifaceted nature of eye activity may complement the similarly multidimensional nature of mental workload.

Mental workload typically refers to the amount of an operator’s limited processing capacity that is required to perform a given task (O’Donnel & Eggemeier, 1986). Initially, mental workload was conceptualised as a unitary construct in which different tasks were allocated capacity from a single, undifferentiated pool. However, it soon became clear that certain types of tasks could be performed simultaneously without any decrement in performance. Theorists such as Wickens (1992) proposed that there are different pools of processing capacity that can be allocated to specific aspects of a task. Processing capacity can be defined along the dimensions of stages of processing, processing codes, and input/output modalities. If an operator performs a task that contains a high
similarity in the capacity demands among the task components, the competition for similar resources will be more severe and a higher workload will result. The implication of this is that different measures of mental workload will be more or less sensitive to the workload demands specific to certain pools of processing capacity. If, for instance, a potential workload hazard is that the system places excessive demands on the operator’s visual input (e.g., by having too many gauges to monitor), the system developer should use a measure that can differentiate between workload demands in the sensory input modalities. On the other hand, if the overall demand of a system is to be measured, the selected measure should be insensitive to the source of workload demands.

The multifaceted nature of mental workload prescribes the future development of mental workload measures. Research should focus on the development of measures that have well-known properties with regards to their sensitivity to the source of workload demands. Of the classes of potential workload measures, including subjective, performance, and analytic, psychophysiological measures seem well suited to encompass the multidimensional nature of mental workload. Physiological measures can be derived from a number of response systems and include heart rate, blood pressure, brain potentials, skin conductance, and respiratory activity. Moreover, even within a particular response system, different response components can be derived. The ocular system is one such example as it can be compartmentalised into the response components of blink duration, blink frequency, fixation duration, fixation frequency, saccadic extent, relative pupil diameter, and blink latency following important task events. These measures are typically derived from recording changes of the cornea-retinal potential that are generated by eye or lid movements via the electrooculogram (EOG) or by using video-scanning techniques. The recent development of remotely positioned face- and eye-tracking devices has also made it possible to acquire measures of
eye activity more unobtrusively than by attaching sensors to the operator as in EOG recordings and with greater accuracy than in video scanning (Pastoor, Lui, & Renault, 1999). Although the measurement of eye activity can require complex recording equipment and signal processing, there are a number of advantages to the method. For instance, the measurement of eye activity is relatively unobtrusive, does not require overt performance on behalf of the operator, measures can provide a continuous assessment to index moment-to-moment changes in workload, and they have the potential to standardise workload levels across different types of tasks or individuals (Kramer, 1991).

Two measures of eye activity that have been applied in previous research include pupil diameter and eye point-of-regard measures. Pupil diameter seems to be correlated with general cognitive demands of a task rather than specific to pools of workload demands. For instance, variations in pupil diameter have been correlated with the comprehension of discrete sentences (Just & Carpenter, 1993) and digit span recall (Granholm, Asarnow, Srkin, & Dykes, 1996). In practical applications, pupil diameter has been shown to increase with increasing visual target density in a mock warfare task (Van Orden et al., 2001). Unlike pupil diameter, eye point-of-regard measures are more specific to visual input demands. Of the eye point-of-regard measures, the frequency of fixation to an object may indicate its importance in a task, whereas fixation duration may indicate the ease in obtaining information from the object. For instance, drivers will show a higher fixation frequency when overtaking bicycles than when overtaking automobiles (Mori, Tanaka, & Abdel-Halim, 1977) and symbolic signs require a shorter fixation duration than alphanumeric signs (Mori & Abdel-Halim, 1981). The comparison of cockpit control panels has shown that some panels require significantly less instrument viewing time by allowing information to be obtained more quickly from the layout and so provide more time for outside-the-cockpit viewing (Gainer &
Obermayer, 1964). Another eye point-of-regard measure, saccadic extent, has been shown to decline with increases in target density during a mock warfare task (Van Orden et al., 2001).

Perhaps the most popular measures of eye activity are linked to endogenous eye blinks. Endogenous eye blinks are spontaneous blinks in the sense that they are not reflexive in response to specific environmental stimuli. Endogenous eye blinks can be measured in terms of blink rate and blink duration. Blink rate generally decreases under conditions of high visual and/or cognitive load such as driving an automobile in the city versus the highway (Pfaff, Fruhstorfer, & Peter, 1976), target detection and weapons delivery in air-to-ground aircraft missions (Wilson, Fullenkamp, & Davis, 1994), and target detection in a mock air warfare task (Van Orden et al., 2001). Since visual input is disrupted when the eye lids close, a reduced blink rate will help maintain continuous visual input when visual workload demands are high. Conversely, people tend to blink more when the eyes move from instrument to instrument because blinks tend to punctuate the end of an episode of information intake (Fogarty & Stern, 1989). Consistent with this interpretation, increased blink rates have been observed with increases in navigational demands of a simulated flight mission (Wierwille, Rahimi, & Casali, 1985). Blink duration is also related to visual workload demands because decreased blink duration will reduce the interruptions of visual input when the eye lids are closed during a blink. Blink duration is reduced in actual versus simulated flight (Wilson, Purvis, Skelly, Fullenkamp, & Davis, 1987) and a dual versus single task condition (Sirevaag, Kramer, deJong, & Mecklinger, 1988).

Eye blinks can also occur following or in response to subthreshold external stimulus events (e.g., following the occurrence of a potential threat target in a radar display). Blinks that occur following the presentation of task-relevant stimuli, but are not elicited reflexively by the physical properties of the stimulus itself, may mark the termination of a stimulus evaluation process (Stern,
Spontaneous and reflexive (Walrath, & Goldstein, 1984). If a motor response must also be performed in reaction to the stimulus, blinks may be further inhibited until the end of the response selection or the motor programming process. Increased workload demands may thus be expected to increase blink latency following the onset of task-relevant events (Stern et al., 1984). Consistent with this notion, the latency of a blink following the occurrence of a task-relevant stimulus has been shown to be longer when the set size is increased in a memory comparison task (Bauer, Goldstein, & Stern, 1987) and when responses are required in auditory discrimination tasks (Goldstein, Walrath, Stern, & Strock, 1985). In comparisons between single-task and dual-task conditions, Sirevaag et al. (1988) showed that blink latency was longer following task-relevant stimuli in the dual-task condition.

Unlike eye blinks that occur following a subthreshold stimulus event, reflexive blinks occur in response to the physical properties of a task-irrelevant sensory stimulus. The blink reflex can be elicited by a sudden and moderately intense stimulus from a range of input modalities, such as a loud noise, flash of light, tap to the forehead, or electrical stimulation of the trigeminal nerve. The most common method of recording blink reflexes is to record the activity of the muscle around the eye that controls blinking, the orbicularis oculi, via electromyographic (EMG) recordings. The typical electrode placement and the resulting raw and integrated EMG recording is shown in Figure 1. However, any measurement technique that can be applied for recording endogenous eye blinks, such as EOG or remotely positioned eye-tracking equipment, may also be used provided that the onset of the blink reflex can be time locked to the eliciting stimulus and that the sampling rate is sufficiently high. Indeed, EOG measures have been successfully applied in the measurement of blink reflexes by several investigators (Hirano, Russell, Ornitz, & Liu, 1996; Simons & Zelson, 1985). Blink magnitude is the most common measure of the blink reflex, although blink latency or response probability can also provide additional information.
Psychophysiologists use the blink reflex to index a variety of perceptual, cognitive, and affective processes (see Dawson, Schell, & Böhmelt, 1999). The blink-eliciting stimulus is used as a probe by presenting it during foreground stimuli and during foreground stimulus-free intervals. For instance, if attentional demands during a warned reaction time task is to be examined, the blink eliciting stimulus can be presented during the warning stimulus (foreground stimulus) and in between successive warning stimulus presentations (foreground stimulus-free interval).

Researchers examine the size of the blink reflex during foreground stimuli relative to the foreground stimulus-free intervals. Blink reflexes elicited during foreground stimuli may thus be facilitated or inhibited relative to blinks elicited during the foreground stimulus-free intervals. The extent of this modulation may also differ between different types of foreground stimuli according to the physical nature of the stimuli or the cognitive processes elicited by the stimuli. As a probe of cognitive processes, the blink reflex offers a number of advantages. The recording of blinks uses a relatively simple procedure that yields a high signal-to-noise ratio, reflexive blinks do not require attention to the eliciting stimulus, and the short latency of blinks (20 to 70 ms) affords good temporal resolution of cognitive events. In addition, because blink reflex modulation can be expressed as a function of the participant’s change from a control or no-task condition, it may provide a standarised measure of attentional demands that is free from individual differences in physiological reactivity. Despite these advantages, the blink reflex has not yet received adequate attention in mental workload research. This is despite laboratory research that has provided a wealth of information about the cognitive variables that affect blink reflex modulation.

Early research that used discrete or short duration foreground stimuli in combination with an independent and well-validated psychophysiological measure of attentional engagement, heart rate deceleration, supported a link between anticipatory attention and blink facilitation (see Graham,
The relationship between the direction of attentional engagement and blink modulation has been examined in experiments that employed foreground stimuli from different sensory modalities during attention demanding tasks. Putnam (1990) conducted experiments in which an acoustic blink-eliciting stimulus was presented during foreground stimuli presented in the acoustic, visual, and tactile modalities. The foreground stimuli signaled various salient events, such as reaction time tasks or perceptual discriminations. The use of foreground stimuli in different modalities with an acoustic blink-eliciting stimulus resulted in conditions where there was a modality match or mismatch. Blinks elicited during the acoustic foreground stimuli (modality match condition) were facilitated relative to baseline. In contrast, blink inhibition was observed when visual and tactile foreground stimuli were used (modality mismatch condition). Moreover, the amount of blink facilitation and inhibition increased towards foreground stimulus offset if the termination of the foreground stimulus conveyed important information (e.g., if it signaled a reaction time response). Modality-specific blink modulation has also been observed when the modality of the blink-eliciting stimulus is varied (Anthony & Graham, 1985). Modality-specific blink modulation has been attributed to the task demands engaging attention towards the sensory modality of the lead stimulus, resulting in an enhancement of other stimuli that are presented in that modality and/or an inhibition of stimuli presented in other modalities (Putnam, 1990; Silverstein, Graham, & Bohlin, 1981).

Modality-specific blink modulation has not been found during some types of tasks. These have included a discrimination and counting task (Lipp, Siddle, & Dall, 1998), an intermodality change task (Lipp, Siddle, & Dall, 2000b), and a reaction time task (Lipp, Siddle, & Dall, 2000a). Although Putnam (1990) reported modality-specific effects in reaction time tasks similar to that used by Lipp et al. (2000a), the procedures may have differed on a number of important dimensions,
such as in the overall task difficulty. Lipp et al. (2000a) suggested that the difficulty of the task may be crucial and that modality-specific effects are observed only during highly engaging tasks that encourage localised orienting, whereas modality nonspecific effects are observed during less engaging tasks that encourage generalised orienting. Neumann and Lipp (2002) examined acoustic blink modulation during a discrimination and counting task under a low and high level of task difficulty. The findings in the low task difficulty condition replicated those reported previously by Lipp et al. (1998). Blink reflexes were facilitated during acoustic and visual foreground stimuli that were attended to relative to foreground stimuli that were ignored. In the high task difficulty condition, blinks were larger during attended stimuli than during ignored stimuli for acoustic foreground stimuli, but not for visual foreground stimuli. These results are thus consistent with the notion that task difficulty may mediate the observation of modality-specific effects. However, blink reflex attenuation was not greater during attended visual foreground stimuli than during ignored visual foreground stimuli in the high task difficulty condition. This suggests that some tasks may require a sufficiently high level of task demands in order to observe modality-specific blink reflex modulation.

A more effective way to increase task engagement may be through the use of continuous task demands. A continuous task will generate a greater cognitive demand than discrete, punctuated tasks due to additional cognitive processes such as response scheduling, monitoring of feedback, and stimulus anticipation. The laboratory-based procedure of the memory-load version of the Continuous Performance Test was used by Hazlett, Dawson, Schell, and Nuechterlein (2001). In this procedure, participants viewed a series of distracter and target digits presented in rapid succession on a computer screen and were required to press a button when the target number 7 appeared immediately after the number 3. Acoustic blinks elicited during the anticipation of a digit
were inhibited relative to baseline. The inhibition did not seem to reflect the effects of processes other than attention, such as arousal or response preparation, because blink inhibition did not differ when the digit anticipation occurred after a distracter digit or after the target digit number 3. Thus, the Hazlett et al. (2001) study provides evidence that modality-specific attentional engagement may be observed during continuous tasks.

The findings using simple laboratory-based procedures provide a strong basis to suggest that the blink reflex will be differentially sensitive to mental workload demands during complex continuous tasks. Moreover, the blink reflex may be sensitive to the workload demands associated with different stimulus input modalities during these tasks. At present, these issues have not been explored in detail. The only study to examine blink modulation in a multiple-task environment was reported by Neumann (in press). In this experiment, participants interacted with a visual display for three minutes under single and multiple task conditions. In the single-task condition, participants performed a horizontal tracking task. In the multiple-task condition, participants performed the tracking task in addition to monitoring the status of four vertical and two dial gauges. The blink reflex to an acoustic reflex stimulus was measured during the tasks to assess the differences to a baseline (no-task) control condition and to assess whether blinks would be differentially sensitive to the workload demands in the single- and multiple-task conditions. Because the workload demands were visual, whereas blink was elicited in the acoustic modality, it was hypothesised that blinks would be inhibited while performing the tasks and that blink inhibition would be greater during the multiple-task condition than during the single-task condition. Convergent validity for the workload manipulations was obtained with subjective ratings on the Modified Cooper-Harper Rating Scale and with additional physiological measures of heart period, heart period variability, and endogenous blink rate. As predicted, blink magnitude was smaller and blink latency was longer when
participants performed the task than when participants simply viewed the movements of the visual display. Moreover, the attenuation of blink magnitude and latency was more pronounced in the multiple-task condition than in the single-task condition.

Two additional issues were examined in Neumann's (in press) experiment. First, it was noted that the presentations of the blink-eliciting stimulus could be intrusive on task performance. The issue of intrusiveness of a workload metric is not unique to blink probe methodology, but is shared by other methods that require the presentation of additional stimuli (e.g., secondary tasks and event-related potentials). The potential intrusiveness of the blink-eliciting stimulus is an important consideration in multiple-task environments as it may interfere with primary task performance and add to the workload. The latter possibility seems unlikely since the blink reflex is a brain-stem reflex that does not require attention to the eliciting stimulus (e.g., it can be elicited during sleep, Bowker & Morrison, 1976). The blink-eliciting stimulus may also disrupt the sensitivity of other concurrently measured workload metrics, particularly other physiological measures, because it causes changes in various physiological systems (e.g., changed heart rate). To examine the issue of intrusiveness, Neumann (in press) required participants to perform the single-task, multiple-task, and baseline conditions for trials that did and did not contain the blink-eliciting stimulus. The presentations of the blink-eliciting stimulus did not impair task performance on the tracking task when performed in isolation or when performed concurrently with the gauge-monitoring task. In addition, all the physiological responses (heart period, endogenous eye blink rate, and heart period variability) were sensitive to the workload demands regardless of whether the blink-eliciting stimulus was presented or not. However, the blink-eliciting stimulus did have a minor effect on heart period variability as assessed by spectral power analysis. During the no-task baseline trials, power in the low and high frequency components was lower when the blink-eliciting stimulus was
The presentations of the blink-eliciting stimulus may have influenced heart period variability by causing a suppression of the cardiovascular control system either directly or indirectly by inducing temporal changes in respiratory activity. For the subjective measures of workload demands, it was found that participants rated the tracking task performed alone as less demanding when the blink-eliciting stimulus was presented than when it was absent. This finding is contrary to expectation, as the presentation of the blink-eliciting stimulus would be expected to increase subjective workload demands. Neumann (in press) suggested that the presentation of the blink-eliciting stimulus may have reduced perceived workload by helping the participant maintain an active attentional set when performing the monotonous tracking task.

Taken together, the findings of Neumann (in press) suggest that the intrusiveness of the blink-eliciting stimulus is limited when used to assess mental workload demands during tasks lasting up to three minutes. Moreover, as the participants repeated the task conditions several times over the 130-minute testing session, fatigue over this time period does not appear to increase the problem of intrusiveness. Independent research also indicates that the intrusiveness of the blink-eliciting stimulus is limited. For instance, Bradley, Cuthbert, and Lang (1990) found that verbal reports of valence, interest, and arousal during emotionally arousing pictures corresponded to facial EMG, heart rate, and skin conductance activity recorded in a separate presentation session that also included presentations of an acoustic blink-eliciting stimulus. Vrana (1995) reported that the interference caused by an acoustic blink-eliciting stimulus on facial EMG recordings was limited to the 0.5 seconds immediately following the stimulus. However, further work is required to examine the issue of intrusiveness of the blink-eliciting stimulus under a wider set of workload conditions. For instance, the blink-eliciting stimulus may be more intrusive on performance when it is presented in the same modality as that which attention is directed or if it is presented during tasks that require
Spontaneous and reflexive a higher level of workload demands or higher precision of motor control. Alternatively, less intrusion may be introduced if the blink-eliciting stimulus is presented via electrical stimulation of the trigeminal nerve or by an air-puff directed below the eye because these methods may not require the eliciting stimulus to be as intense as that used with acoustic methods of blink elicitation.

The second issue examined by Neumann (in press) concerned the habituation of the blink reflex. The blink reflex will attenuate with repeated presentations of the eliciting stimulus. Habituation has been shown to influence blink modulation when assessed during a prepulse inhibition paradigm (Lipp & Krinitzky, 1998). In the prepulse inhibition paradigm, the blink-eliciting stimulus is presented approximately 100 ms following the onset of a foreground stimulus. To examine whether habituation influenced the modulation of blinks during the tasks, Neumann (in press) measured blink reflexes in baseline trials before and after the experimental session and repeated the single- and multiple-task trials across the session. The findings yielded a number of interesting interactions between task condition and the measure of the blink reflex. Habituation was evident in blink magnitude, but not blink latency, when the pre- and post-session baseline trials were compared. During the single- and multiple-task trials, habituation was evident in both measures. However, habituation of magnitude interacted with task condition in that blink magnitude habituated during the single-task condition, but not during the multiple-task condition. It may be that the increased motor demands in the multiple-task condition reduced the development of habituation relative to the single-task condition. Similar reduced rates of habituation have been found in other research when participants prepare to execute a motor response (Valls-Sole, Valldeoriola, Tolosa, & Nobbe, 1997). However, even though habituation of blink reflexes was evident across the experimental session, this did not affect the sensitivity of the measure to detecting differences in the workload demands.
The study by Neumann (in press) has provided some promising, yet preliminary findings indicating that the blink reflex is sensitive to mental workload demands in complex control tasks. O’Donnel and Eggemeier (1986) proposed several properties that should be considered when selecting workload metrics for a particular application. These properties of sensitivity, diagnosticticity, intrusiveness, validity, reliability, operator acceptance, and ease of application, can help structure future research on the utility of blink reflexes as a measure of mental workload. Neumann’s (in press) study has already shown that blink reflexes are sensitive to the workload demands in a single-task versus a multiple-task comparison. However, it is not clear whether blink reflexes will also prove sensitive to various levels of task demands within a multiple task environment or will be sensitive to changes in workload demands across training in a task. The blink reflex may also be sensitive to workload demands that follow imperative events in the environment (e.g., error signals in the system). The high temporal resolution of blink reflexes may enable the examination of workload demands during components of a task. A possible advantage of blink reflexes in workload measurement is that it may prove diagnostic of the mental workload demands in different input modalities. Future research could use different combinations of modality tasks and vary the workload demands on one stimulus input modality to determine whether the diagnosticticity for input modality remains in multimodal environments. The intrusiveness of blink reflex measurement appears to have been minor in the preparation used by Neumann (in press). It remains to be seen whether the intrusiveness is also negligible in other types of multiple tasks. Validity and reliability are important features that must be confirmed in any workload metric to be sure that the methodology actually does measure the intended construct and can do it consistently over time. The systematic evaluation of these properties will help determine the utility of the blink reflex as a workload metric, at least in comparison to other measures. The
application of blink reflexes for the development and testing of complex control systems in the field will also be reliant on studies which show that the method is well accepted by the operator and can be easily applied. If operator acceptance is low, it may reduce the potential usefulness of the measure in the field.

To summarise, if measurements of eye activity such as blink reflexes are to gain acceptance as workload metrics, they must prove to be valid and reliable measures of mental workload. The measures must also prove to have unique advantages over other metrics such as by enabling the measurement of aspects of mental workload demands not otherwise possible, by enabling the measurement of mental workload in situations not otherwise possible, or by being easier to apply in practice than other measures. A complete assessment of eye activity has the potential to aid an investigator to diagnose problems in existing systems or to evaluate the validity of changes in system design that aim to reduce workload. For instance, pupil diameter would provide a global measure of workload demands, whereas endogenous eye blinks would be more diagnostic of workload demands in the visual input modality. The efficiency of information retrieval in a system may be derived from the latency of eye blinks following task-relevant events or from examining patterns in eye point-of-regard measures. Alternatively, modulation of the blink reflex could be measured if the level of mental workload during specific stimulus events is to be investigated.
References


Figure

Figure 1. The typical electrode placement and the resulting raw and integrated signals that are obtained during electromyographic recordings of the eye blink reflex.
Rectified and integrated blink reflex response

Raw electromyographic recording