Pupil Dilation and Eye-tracking

Joseph Tao-yi Wang[*]

I. Introduction

A video-based eye-tracker (e.g., the mobile Eyelink II or Eyelink Remote of SR research, Osgoode, Ontario, Canada) uses video cameras to record the eye position of human subjects, and hence, record pupil dilation and eye movements. The eyetracker puts a video cameras and infrared illuminators in front of the eye to record the position of the eye, cornea reflections, and the size of the pupil. Using the movements of one's pupil (with respect to cornea reflections), the eye-tracker tracks the movement of one's eyes, which is then mapped into locations on the screen by calibration and adjustments for head movements.

With the eyetracker, we can measure gaze locations, the time length of fixations, and pupil dilation. Hence, using the eyetracker, we can investigate how fixations (looking at the same place for a while), saccades (fast eye movements) and pupil dilation responses (changes in pupil sizes) are related to the information on the screen and behavioral choices during an experiment. Understanding the relationship between these observables can help us understand how human behavior in the economy can be affected by what information people acquire, where their attention is focused on, what emotional state they are in, and even what brain activity they are engaged in. This is because fixations and saccades (matched with information shown on screen) indicate how people acquire information (and what they see), time lengths of fixations indicate attention, and pupil dilation responses indicate emotion, arousal, stress, pain, or cognitive load.

---

[*] October 20, 2009. Department of Economics, National Taiwan University, 21 Hsu-Chow Road, Taipei 100, Taiwan (email: josephw@ntu.edu.tw). This is a book chapter in preparation for the *Handbook of Process-Tracing Methods,* to be published by the *Psychology Press.* Please do not quote without permission by the author.

This chapter attempts to demonstrate how these measures in pupil dilation and eye-tracking can help to study human decision making processes. In section II, we introduce video-based eye-tracking. Section III reviews the literature of pupil dilation responses and discusses issues regarding interpretation. Section IV discusses applications of pupil dilation on various tasks of decision making. Section V concludes with a list for further reading. There is also an appendix documenting the "standard operational procedure" of performing eye-tracking studies using either the Eyelink II or Eyelink Remote eye-tracker, intended as a guide for readers who are interested in conducting their own eye-tracking research.

II. Eye-Tracking and Measuring Pupil Dilation: How it works

There are several different ways to track eye movements and measure pupil dilation. See Duchowski (2007) for a complete review of all the different techniques. One of the most reliable non-invasive methods utilizes high speed video cameras. In particular, video-based eye-trackers put cameras and infrared illuminators in front of subject's eyes, and videotape eye movements and corneal reflections. This is typically performed by either placing cameras in front of the computer screen subjects are viewing (desk-mount), or by placing cameras on a head restraint similar to a bicycle helmet so they are located right in front of each eye (head-mount). Since images of the pupil are recorded, the eye-tracker is able to measure pupil dilation by either counting the number of pixels of the (dark color) pupillary area or fitting an ellipse on the pupil image and calculating the length of the major axis.[1] Moreover, using the movements of subjects'

---

[1] See Klingner et al (2008) for a comparison of the accuracy of the two methods. Note that since images of the pupil would vary in size depending on camera position, pupil sizes are measured only relatively, unless one provides a benchmark to compare with. This could be a problem if one needs absolute measures of pupil size, which is more reliable. In fact, as reviewed by Beatty and Locero-Wagoner (2000), changes in absolute pupillary diameters are robust to baseline pupil size, say, due to changes in luminance.

pupil relative to the cornea reflection, the eyetracker tracks the movement of their eyes. A "calibration" procedure is performed by showing black dots at several baseline positions on the screen (such as corners, center, etc.), and having the subject gaze at them. Typically, a nine point calibration is sufficient. If the eye-tracker is head-mounted, further adjustments for head movements are automated by tracing sensors on the four edges of the screen. If the eye-tracker is remotely positioned (typically in front of the screen), an additional target on the forehead is used.[2] When compared to images when subjects were requested to fixate on several baseline positions on the screen, the eye-tracker can interpolate and infer the current location the subject is looking at on the screen for each snap shot in the video. This allows the experimenter to trace eye movements on the screen and infer lookup patterns.

Moreover, since we record gaze locations almost in real time, we can infer the time length of these fixations. However, the accuracy of the length of the fixations depends on the sampling rate, or how frequent gaze locations are recorded. For example, the mobile Eyelink II eyetracker (SR research, Osgoode, Ontario) can sample both eyes at 250Hz (every 4ms). Hence, if the eyetracker records a subject looking at the same location 30 times in a row, we can infer the fixation was 120ms, with a maximum error margin of 4ms. On the other hand, the more advanced Eyelink 2000 (SR research, Osgoode, Ontario) can sample a single eye at 2000Hz (every 0.5ms), or both eyes at 1000Hz (every 1ms), lowering the error margin down to less than 1ms.

---

[2] Other possible head-correcting methods including the use of a chin rest, and software corrections. For example, the Eyelink 2000 eye-tracker uses a chin rest to avoid head movements during high speed tracking. The Tobii eye-tracker claims to use sophisticated algorithms to calculate relative head position. This requires extensive computational power provided by the eye-tracker host computer. Hence, the Tobii eye-tracker has a sampling rate around 50-150Hz.

Since the technology behind the eye-tracker is straightforward, amateurs could in principle build their own eye-trackers even using video camcorders.[3] However, researchers should look for commercialized eye-trackers as they are more advanced and provide high-speed recording (currently up to 2000Hz, or 0.5ms per image, up from 250-500Hz a few years ago) and more accurate eyetracking results (tracking errors typically lower than 0.5 degrees). These advanced eye-trackers are also equipped with software for data analysis, which can process the raw data and output them in various forms, ranging from different "reports" in plain text format (to be imported into any statistical software), figures (such as heat maps, fixation maps and saccade maps), to animated video replay of the eye movement sequence, and come with their own experimental software, or supporting drivers for existing software such as E-prime and Matlab's psychophysics toolbox, which are essential to conducting the experiment. Though most researchers cannot easily amass dozens of eyetrackers to study asset market behavior or strategic interactions in large groups, in the realm of small scale experiments, such as individual decision making, paired experiments, and most pilot studies, commercialized eye-trackers can be utilized to obtain reliable eye-tracking measures while subjects perform the behavioral task. Hybrid experiments that combine commercialized eye-trackers with other non-tracking computer terminals are also possible, and could be a good compromise when the number of eye-trackers is limited.[4]

Video-based eye-trackers report raw data (usually called the "sample report") for each instance of observation. Each record consists of a time stamp, the X-Y location of the left and

---

[3] For example, if you perform a Google search on keywords such as "building an eye tracker", the first two search results are the paper http://www.cis.rit.edu/people/faculty/pelz/publications/ETRA04_babcock_pelz.pdf and the site http://www.cogain.org/eyetrackers/low-cost-eye-trackers, both providing tips to build homemade eye-trackers.

[4] The hybrid approach has one caveat when studying individual differences in the lookup patterns: the eye-tracked subjects may or may not be the particular "type" of subjects one is interested in. Running many sessions or pre-screening for specific types (using a behavioral experiment) would be needed in this case.

right eye (or just one eye under monocular mode), the pupil dilation of the eye(s), and possible "messages" that are sent to the eye-tracker. This provides the basis to calculate more sophisticated statistics, such as velocity, acceleration, saccades, and fixations. Researchers usually focus on the number of fixations and total dwell time of specific "regions of interest (ROI)." Details of these eye-tracking measures can be found in chapter 2.

Note that there are previous "eye-tracking" studies that used mouse-tracking systems (e.g. "Mouselab") to record mouse movements, such as moving a cursor into a box or clicking on a box opens its contents. These "eye movements", or more accurately, mouse movements, could also measure lookup counts and duration. See for example, Camerer, Johnson, Rymon, and Sen (1993); Costa-Gomes, Crawford, and Broseta (2001); Johnson et al. (2002); Costa-Gomes and Crawford (2006); Gabaix, Laibson, Moloche, and Weinberg (2006); and Crawford (2008). One small defect of this system is that the experimenter cannot be certain the subject is actually looking at (and processing) the contents of the open box. After all, Mouselab uses a mouse-tracking technology to proxy for eye-tracking. In contrast, video-based eye-tracking systems measure eye locations so we can tell if the subject's eye is wandering, and pupil dilation is measured at the same time (which Mouselab cannot do).[5]

Lohse and Johnson (1996) compared mouse-tracking with eye-tracking, and concluded that subjects adopt specific information acquisition techniques to deal with the increased search cost caused by the mouse. Hence, instead of using the same eye-tracking measures (such as those discussed in chapter 2), Costa-Gomes et al. (2001) proposed the "Occurrence" and "Adjacency" measure. Later, Costa-Gomes and Crawford (2006) used the "Compliance" rates mainly based on "Adjacency". Note that "Occurrence" does capture similar ideas present in the

---

[5] However, Mouselab systems can be installed cheaply in many computers to measure lookups of many subjects all at once. This is useful in running "efficient" sessions (instead of "wasting" the untracked subjects each time) and studying attention simultaneously in large market experiments.

eye-tracking measure of fixations, while the “Adjacency” measure considers how close certain combinations of lookups are present in lookup sequences and partly captures the order of information acquisition.

III. Pupil Dilation Responses

Human pupils may dilate for various reasons, including memory load, cognitive difficulty, valence, arousal, pain, and so on (Beatty, 1982). For single responses, Hess (1972) reported that such dilation would occur 2-7 seconds after emotional stimuli were presented and suggested faster dilation for stronger stimuli. Partala and Surakka (2003) showed that there was first seemingly no response for about 400ms, and then a steep increase in pupil size peaking at 2-3 seconds after stimulus onset, using different sounds (baby crying, laugh or regular office noise). When performing a cognitively demanding processes, pupils dilate in responses to mental workload of the task and peaks at about 1-2 seconds after the onset of demand (Beatty, 1982), and constrict after the task is completed, either gradually (Kahneman and Beatty, 1966; Hess, 1972) or instantly (Bernhardt et al., 1996).

Beatty (1982) reviewed relevant literature and concluded that task-evoked pupillary responses appeared to be a consistent index of the cognitive load within tasks, across tasks, and possibly even across individuals. Hence, Beatty and Lucero-Wagoner (2000) suggested three measures to report for each time interval of interest: Mean pupil dilation, peak dilation (especially when mean pupil dilation is biased due to varying time length across subjects or overlapping intervals of interest) and latency to peak. Researchers then compared these measures to an established baseline (say, by fixating at a blank screen for several seconds) and report the difference (either in percentages or absolute measures). Beatty and Lucero-Wagoner

(2000) noted that percentage measures are inflated when baseline pupil size is small (due to luminance), and hence, recommended the absolute measure (in mm).

In addition, since pupil dilation is measured continuously throughout the task and has low latency (0.1-0.5 seconds) with respect to changes in mental workload, Kramer (1991) argued that it could provide a more reliable measure for processing demand in general, compared other measures such as event-related brain potential (ERP), electro-encephalographic activity (EEG), and so on. Bailey and Iqbal (2008) utilized pupillary responses to provide a steady stream of workload data to study its changes within a single task. In particular, Bailey and Iqbal (2008) asked subjects to perform three different tasks, route-planning, document editing, and email classification, and decomposed the tasks into various subtasks each assigned to a particular hierarchical level. They then used the average percentage change in pupil size (APCPS) to measure the workload of each subtask (to identify the best timing for computer interruptions).

III.A What Affects Pupillary Responses?

Regarding pupil dilation as an emotional reaction, Hess and Plott (1960) reported pupillary dilation responses to what they call "emotionally toned or interesting visual stimuli". Other early studies reporting pupillary responses to specifically sexual arousal include Hicks, Reaney, and Hill (1967), and Bull and Shead (1979). More recently, Aboyoun and Dabbs (1998) also reported pupillary responses to arousal. Chapman et al. (1999) found that pupil dilation responses to pain began at 0.33 seconds and peaked at 1.25 seconds after stimulus onset. Peak dilation increased significantly as pain intensity increased. Oka et al. (2000) reported similar pupillary responses when pain was administered to either finger tips and ear lobes, and found only minor gender (women had borderline larger peak dilation and faster recovery latency) and age (older subjects had larger delays) differences. Portala and Surakka (2003) reported larger mean pupil diameter (0.2mm vs. 0.14mm) when subjects listened to affect sounds, both positive (baby laughing) and negative (baby crying), compared to neutral sounds (office noise). Pupillary responses to positive sounds were very similar to negative ones, though female responded stronger positive sounds and male stronger to negative ones. Regarding task-evoked pupillary responses to cognitive load, Hess and Plot (1964) first reported differential pupillary dilation responses while mentally calculating the product of two numbers. In particular, pupil dilation increased about twice as much (22 per cent vs. 11 per cent) when subject calculated 16 times 23, compared to 7 times 8. Kahneman and Beatty (1966) showed how more difficult memory tasks (memorizing numbers with more digits vs. less digits) induced larger pupillary response (0.1mm vs. 0.55mm for 3 vs. 7 digits), establishing the link between pupil dilation and memory load. Later studies establish pupillary dilation responses as an indicator of cognitive load in various tasks, including signal detection and letter matching. Goldwater (1972) and Beatty (1982) reviewed the literature on pupillary response as an indicator of cognitive load.

Hence, pupillary responses could be used to measure differences in cognitive load under various tasks, and is applied to study various human behavior such as language processing, and curiosity. For example, Just and Carpenter (1993) used pupillary dilation responses as an indicator of cognitive load during syntactic processing. In particular, they gave subjects object-relative sentences (“The reporter that the senator attacked admitted the error”) that imposes a larger load on short-term memory, and the less cognitive demanding subject-relative sentences (“The reporter that attacked the senator admitted the error”), and asked subject true-false questions later to test their comprehension of the sentences. Object relative sentences induced larger pupillary responses (0.25mm vs. 0.21mm) and increased latency to peak by 116ms. Hyona, Tommola, and Alaja (1995) reported differential mean pupillary diameters when Finnish subjects were listening to a foreign language (4.20mm), shadowing a foreign language (4.72mm), and simultaneously translating a foreign language into their native language (5.22mm). More recently, Kang et al. (2009) showed that pupils dilate in anticipation of seeing the answers to trivia questions that people report they are curious about, and pupil dilation is larger if they were more curious about the answer (8% vs. 4% for high vs. low curiosity).

III.B Interpretations of Pupillary Responses

Since there are various causes that could trigger pupillary responses, care must be taken to distinguish the exact cause that activated the response. (Note that a similar "reverse inference" occurs when interpreting fMRI data, as most brain regions are typically involved in multiple functions.) Here we discuss some of these issues and possible solutions.

The first cause of pupillary response one needs to rule out is pupillary light reflex. This is typically done by controlling luminance during the course of the experiment. For example, Beatty (1972) contains instructions on producing stimulus slides with comparable brightness and contrast. This is more easily done with computerized display, but still remains a challenge in many naturally-occurring settings, such as flying an airplane at night or viewing online search results that contain images of different colors. For example, Dehais et al. (2008) embedded an eye-tracker in a real aircraft to record pupillary responses (as well as fixations) when pilots perform different flying routines, but eventually decided to analyze pupil dilation for only a subset of the pilots to make luminance conditions comparable.

Researchers have developed several statistical treatments to remove the influence due to changes in luminance. For instance, when studying the pupillary response to viewing relevant web search results, Oliveira et al. (2009) performed principle component analysis on the pupil diameter data explicitly to separate the effect of changes in luminance (due to color difference of the results of a Google Image search) from stimulus relevance. On the other hand, since the pupil's response to luminance is either rapid constriction (light) or slow dilation (darkness), while effortful cognitive processing triggers small but rapid increases in pupil size, Marshall (2007) proposed an "index of cognitive activity" based on wavelet theory that captures these unusual increases, making the index insensitive to changes in light.

Another measurement issue researchers have to deal with is possible measurement errors due to gaze angle. In particular, Pomplun and Sunkara (2003) reported distortion to pupil size data reported by video-based eye-trackers, but also proposed a neural-network based calibration interface to correct it. Klingner et al. (2008) reported that using an ellipse-fitting method to measure pupil size could avoid this problem (as opposed to the pixel-counting method used by Pomplun and Sunkara, 2003).

The main issue regarding interpretation is to identify the exact cause of the pupil dilation. For studies conducted to observe pupillary responses to specific tasks, this is done by designing control trials that is identical to the treatment trials except for only one particular factor of interest. For example, to identify the pupillary responses to (positive) affect, Partala and Surakka (2003) compared mean pupil size of trials where subjects heard baby laughing (positive affect) with that of trials where subjects heard regular office noise (neutral). Oliveira et al. (2009) asked subjects to investigate three different Google search results sequentially, and determine which was most relevant. They found pupillary responses to the relevance of search results when the results were each shown for five seconds. However, they also noted that their analysis was made possible by separating the search process into several stages, which is typically not the case in actual web searches.

In more naturally-occurring settings, pupillary responses could be compared with other subjective or psychophysical measures, such as skin conductance, heart rate variability, blinks, subjective ratings, etc. By combining the results from various measures, researchers are more likely to identify the unique cause that could explain all of them. Early attempts in combining pupil dilation and other techniques include Kahneman, Tursky, Shapiro, and Crider (1969) and Just, Carpenter, Keller, Eddy, and Thulborn (1996). Kahneman et al. (1969) combined pupillary responses with skin conductance and heart rate, while Just et al. (1996) performed the same language task as Just et al. (1993), but under positron emission tomography (PET).

More recently, Lin et al. (2007) and Lin et al. (2008) compared pupillary responses and heart rate variability (as well as subjective techniques) to assess user cost of playing a computer game. De Greef et al. (2009), Dehais et al. (2008), Recarte et al. (2008), and Marshall (2007) compared pupil dilation with other eye-tracking measures such as blink rates, fixation time, saccade distance and speed, under various tasks ranging from piloting an airplane, operating a navel warship, driving a car, to more abstract tasks such as problem-solving and visual search. Marshall (2007) concluded a combination of seven “eye metrics” could successfully identify subject’s cognitive state.

In some studies, pupil dilation is used to inform interpretation and analysis of functional magnetic resonance imaging (fMRI) data. For example, Siegle et al. (2003) compared the time course of pupil dilation with that of the fMRI signal in the middle frontal gyrus during a digit sorting task to suggest that activity in that area indexed the working memory subtask of digit sorting. Gilzenrat et al. (2003) reported that pupil diameter correlates well with locus coeruleus (LC) mediated task engagement behavior, and is used to support theories concerning the function of the LC tonic mode in Aston-Jones et al. (2005).

In other studies, pupil dilation is aided by other measurements to provide more accurate interpretation. In Kang et al. (2009), the pupillary response before seeing the answer to an interesting question could be attributed to either arousal due to anticipation of the answer, or pain due to the impatience waiting for the answer, but was interpreted as "anticipation of an interesting answer" because the same self-reported curiosity is also shown by fMRI to activate the ventral striatum, a region involved in anticipated reward.[6]

Similarly, in Wang et al. (2009), pupil dilation is proportional to the size of the lie (how much subjects inflate the true state), which could be due to simply guilt, or a more complicated process involved in lying. The guilt story was suggested by experiments reported in Gneezy (2005) and Hurkens and Kartik (2009), and modeled theoretically by Kartik, Ottaviani, and Squintani (2007), Chen (2007) and Kartik (2008), while the cognitive difficulty story was modeled and experimentally tested by Cai and Wang (2006) and Sánchez-Pagés and Vorsatz (2007). Since their experiments were not designed to distinguish the two stories, the pupillary response results were consistent with both accounts. In fact, Wang et al. (2009) eventually affirmed the cognitive difficulty story more than the guilt one because they found individual differences in subject behavior which could be linked to their lookup patterns, and both results could be explained by a bounded rationality model (the level-k model with heterogeneous types each performing different steps of thinking).

Specifically, the strategic information transmission game studied in Wang et al. (2009) consists of an informed sender (e.g. stock analyst) that sends a (possibly deceptive) message to an uninformed receiver who will then take action (e.g. invest in the stock). The level-k model for this sender-receiver game predicted there would be level-0 senders who always report the

---

[6] Another piece of evidence is the time series of pupil size, which gradually increase during the count down and peaks immediately after the answer is displayed. This favors the anticipation story since impatience should decrease rather than increase as the remaining waiting time decreases.

true state of the world, level-0 receivers who naively follow advice of the sender, level-1 senders who take advantage of level-0 receivers and report their preferred state (send an exaggerated message), level-1 receivers who discount sender messages (anticipating sender's exaggeration), and so on. These level-k types would each predict particular behavior, specific steps of reasoning (best responding to opponents down a hierarchy), and lookup patterns that were matched with individual-level data.

Note that the level-k model predicts specific steps of reason for each type (the "best response hierarchy"), and provides a particular decomposition of the decision making process which could be empirically tested (with either behavioral or psychophysical data). In fact, researchers in computer science have been performing similar decomposition of tasks using well-known modeling techniques such as GOMS (Card et al., 1983, John, 1995, John and Kieras, 1996). These models of task execution also played important roles in explaining pupillary responses in human-computer interaction (HCI). For example, Iqbal et al. (2004) recorded pupillary responses when subjects were performing reading comprehension, mathematical reasoning, product searching, and email classification. Lin et al. (2007, 2008) observed pupillary responses when subjects played a video game. Both found little difference in pupil dilation for simple and difficult tasks when averaged across entire trials, but Iqbal et al. (2004) did find significant difference in "cognitive subtasks" (but not in "motor subtasks") when they decomposed the original task into several lower-level subtasks using GOMS analysis. Bailey et al. (2007) developed publicly available software (TAPRAV) to perform similar task decomposition and link it with pupil dilation data.

There are also task decomposition techniques based on an entire sequence of fixations over the period of decision-making (which represents a particular order of information acquisition or attention during that period). These models could also be utilized to analyze pupillary responses and eye-tracking data. In particular, with predefined (economic) models about the decision-making process, researchers could generate a certain sequence of lookup behavior, and test if an empirical sequence of fixations is consistent with the particular reasoning process derived by the model. For example, Hunton and McEwen (1997) modeled information search strategies of financial analysts and studied their effect on forecasting accuracy utilizing an eye-tracking technology developed for disable analysts. Also, researchers have been using fixation sequences to study human read (Rayner, 1998). To formulate such assumptions, we require a predefined model about the decision-making process that generates specific predictions regarding the sequence of fixations.[7]

Since such a task is context specific, we illustrate this method by an example: Chen et al. (2009) considered a two-player spatial location guessing game in which each side chooses a location *(x,y)* on a commonly known N-by-M map, attempting to fulfill one's goal of being located close to a certain target based on the opponent's location, such as "two steps to the right of your opponent", namely (2,0), and "three steps step above your opponent", or (0,3). Existing literature provide a "level-k" model to explain initial responses when subjects first see this game (cf. Costa-Gomes & Crawford, 2006). Specifically, subjects behave heterogeneously, and could be classified as various level-k types, each holding different beliefs and exhibiting different degrees of sophistication. The naïve L0 type subjects simply pick the center, namely (0,0), or randomly (so the average would be the center). Believing that they are facing L0 opponents, L1 type subjects choose the location "two steps to the right of the center", namely (2,0), and "three steps above the center", i.e. (0,3). Believing they are facing L1 opponents, L2 type subjects

---

[7] Note that the mere existence of such models could be an issue. In fact, there is no model that is universally agreed upon in reading studies.

choose the location “two steps to the right of the location which is three steps above the center” and “three steps above the location which is two steps to the right of the center”, both meaning (2,3). Higher types are defined iteratively until they hit the boundary of the map, in which they coincide with equilibrium. Figure 1 illustrates equilibrium and level-k predictions for a game with N=M=9 and goals (2,0) and (0,3).

This model simultaneously provides a specific algorithm to find the choice of a particular level-k type, which consists of a plausible sequence of locations the same level-k type subject would look at on the screen. For example, the L2 types would choose the location (2,3), and look at {(0,0), (2, 0), (3,2)} or {(0,0), (0, 3), (3,2), depending on which goal they have. This sequence could be captured through eye-tracking, and distinguished from, say the lookup predictions of L1 types, namely {(0,0), (0,3)} or {(0,0), (2,0)}. In fact, Chen et al. (2009) estimated a state-switching lookup model for subjects’ lookup sequences modeling the change in the (unobservable) reasoning states (L0, L1, etc.) as a Markov chain, and the lookup location conditional on each state as a logit distribution centered at the predicted state location.

IV. Applications in Decision-making

Since pupillary responses are a reliable measure of workload and affect, it can be used in various fields to answer different questions regarding decision-making.

First of all, as discussed above, researchers in computer science have been studying human-computer interaction (HCI) by using pupil dilation as an indicator for workload (Iqbal et al., 2004, Lin et al., 2007, 2008, Bailey and Iqbal, 2008, Oliveira et al., 2009). Their results show that although pupillary responses may not be significantly correlated with workload when averaged over the entire task (Iqbal et al., 2004, Lin et al., 2007, 2008), one could still observe correlation between task difficulty and pupil dilation if they focused on specific stages of the task (Iqbal et al., 2004, Bailey and Iqbal, 2008, Oliveira et al., 2009). Hence, researchers could perform similar exercises to analyze how mental workload differs across stages of subject's decision-making process. This would be particular useful if the process consists of dual systems (such as automated vs. controlled) that interact, as is the case of many stroop tasks.

Other decision processes such as learning, which consists of several stages (such as information acquisition, cognitive reasoning, and then action), could also be analyzed. In fact, researchers in education have been using pupillary responses as one of the measurements for "cognitive load." See Paas et al. (2003) for a review of how these measurements contribute to cognitive load theory (CLT) in education. This has been combined with HCI research to study e-learning. For example, in a computer-based adaptive learning environment, Muldner et al. (2009) decomposed subject's learning process by coding their reasoning behavior (self-explanation, analogy, and other) and affect (positive, negative). They found significantly larger pupil size when subjects experienced positive affect compared to negative ones. This suggests the learning process involved stronger excitement of learning compared to the frustration of learning, which indicates positive e-learning experience (though further investigation is needed). They also found significantly larger pupil size when subjects were self-explaining compared to "other reasoning," while the difference between self-explanaion and analogy was insignificant. This indicates self-explanation induces a significantly higher workload than "other reasoning," but not "analogy." Conati and Merten (2007) also found positive but insignificant increases in pupillary response when subjects were self-explaining (compared to not).

Moreover, researchers in e-commerce have attempted to utilize eye-tracking to build systems that recognize consumer's attention and affect regarding certain products and respond by offering that particular product. In particular, Bee et al. (2006) built an “AutoSelect” system that selects one necktie out of a pair of ties presented on screen and asks subjects to verify if this is what they prefer. This system was mainly based on gaze bias studied by Shimojo et al. (2003), and achieved 81% accuracy rate in selecting subject’s preferred choice. Other affect-related signals, such as skin conductance, blood volume pulse, eye blink rates, pupil size, etc., could also be incorporated into the system.

On the other hand, pupil dilation during deception was discovered and studied as early as Berrien and Huntington (1942), though the psychophysical pathways underlying deception might not been fully understood in that time. Various later attempts of using pupillary response as a lie-detecting device include Heilveil (1976), Janisse (1973), Bradley and Janisse (1979, 1981), Janisse and Bradley (1980), Lubow and Fein (1996). These studies found increased pupil dilation when subjects gave deception answers to demographic questions, respond to questionin of “guilty information” or view photos that contain items present at the crime scene. Dionisio et al. (2001) asked subjects to answer questions regarding memory of general knowledge (“What are the colors of the American flag?”) or specific scenarios (“What was the name of the person in the story?”), and compared pupillary responses when subjects were instructed to generate false answers or answer accurately. They specifically interpreted pupillary dilation as an index of cognitive load and concluded that “try[ing] to make their lies as believable as possible” was a more cognitively demanding task than truth-telling.

There are two issues that deserve further investigation. Most lie-detection studies utilized mock crimes (or true and false statements) and simply instructed them to either lie or tell the truth, no serious consequences were actually imposed.[8] But even in studies that do provide

---

[8] This is common to most lie-detection studies. For example, Spence et al. (2001) utilized fMRI to study neural correlates of deception, but all subjects in were told that they successfully fooled the investigators who tried to

monetary incentives to induce naturally-occurring lies, such as the strategic information transmission game (Crawford & Sobel, 1982) studied by Wang et al. (2009), in which an informed sender (e.g. stock analyst) sends a (possibly deceptive) message to a receiver who will then take action (e.g. invest in the stock), researchers still finds pupillary responses to deception. Hence, further research is required to determine the extent hypothetical inquiries affect pupillary responses to deceptions. The possibility of counter-measures is another concern regarding using pupillary responses as lie-detectors. For example, Ekman et al. (2008a, b) discussed the possibility of training subjects to provide voluntary pupillary responses, and even designed a computer game that utilizes pupillary responses as one of the inputs. Hence, whether people could be trained to produce certain pupillary responses to avoid detection when lying demands further research to clarify.

Finally, pupillary responses could be applied to real world scenarios and measure workload and/or affect induced by specific tasks. This information could then be used to understand the decision-making process in more practical settings (compared to controlled environments). Despite various obstacles (especially on mobility issues of the apparatus), there are several successful attempts. For instance, Recarte and Nunes (2003) investigated the effect of mental workload (manipulated by multi-tasking) when driving and found pupillary responses were still consistent with other measures of workload in such settings. Dehais et al. (2008) showed pupillary responses when airplane pilots were performing an engine failure exercise despite luminosity issues. In an abstracted version of the combat management workstation

---

detect them. Moreover, in Frank and Ekman (1997), subjects were threatened with “sitting on a cold, metal chair inside a cramped, darkened room labeled ominously XXX, where they would have to endure anywhere from 10 to 40 randomly sequenced, 110-decibel startling blasts of white noise over the course of 1 hr”. However, this punishment was never actually enforced. If no subject is actually caught or punished, one wonders the credibility of such punishments.

abroad naval vessels, De Greef et al. (2009) found pupillary responses in the “overload” and “normal” scenarios in comparison with the “under-load” scenario.

V. Recommended List for Further Reading:

The handbook chapter by Beatty and Lucero-Wagoner (2000) is a good source for basic knowledge in the psychophysiology of pupil dilation. In this chapter, the authors reviewed the biological foundation of pupil dilation, and provided a thorough review on the literature of “cognitive pupillometry”, or using pupillary dilation responses as an index of brain activity. Duchowski (2007)’s revised book documents the methodology of eye-tracking in general. For those who are interested in exploring workload differences in different stages of decision-making and align pupillary responses with models of task execution, Bailey et al. (2007) presents a publicly available software TAPRAV. If one is interested in conducting similar research, carefully evaluating one of the recent studies employing pupil dilation and eye-tracking, such as those introduced in the previous section, would provide a practical guide to perform it and provide examples to what could be done.

References

Aboyoun, D. C. & Dabbs, J. N. (1998). The Hess Pupil Dilation Findings: Sex or Novelty? *Social Behavior and Personality, 26*(4), 415-419.

Aston-Jones, G. & Cohen, J. D. (2005). AN INTEGRATIVE THEORY OF LOCUS COERULEUS-NOREPINEPHRINE FUNCTION: Adaptive Gain and Optimal Performance. *Annual Review of Neuroscience, 28*(1), 403-450.

Bailey, B. P., Busbey, C. W., & Iqbal, S. T. (2007). TAPRAV: An interactive analysis tool for exploring workload aligned to models of task execution. *Interacting with Computers, 19*(3), 314-329.

Bailey, B. P. & Iqbal, S. T. (2008). Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM Trans. Comput.-Hum. Interact., 14*(4), 1-28.

Beatty, J. (1982). Task-Evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources. *Psychological Bulletin, 91*(2), 276-292.

Beatty, J. & Lucero-Wagoner, B. (2000). The Pupillary System. In J. T. Cacioppo, L. G. Tassinary & G. G. Berntson (Eds.), *Handbook of Psychophysiology* (2nd edition ed., pp. 142-162). Cambridge: Cambridge University Press.

Bee, N., Prendinger, H., Nakasone, A., André, E., & Ishizuka, M. (2006). AutoSelect: What You Want Is What You Get: Real-Time Processing of Visual Attention and Affect *Perception and Interactive Technologies* (pp. 40-52): Springer-Verlag.

Berrien, F. K. & Huntington, G. H. (1943). An exploratory study of pupillary responses during deception. *Journal of Experimental Psychology, 32*(5), 443-449.

Bradley, M. T. & Janisse, M. P. (1979). Pupil Size and Lie Detection - the Effect of Certainty on

Detection. *Psychology, 16*(4), 33-39.

Bradley, M. T. & Janisse, M. P. (1981). Accuracy Demonstrations, Threat, and the Detection of Deception - Cardiovascular, Electrodermal, and Pupillary Measures. *Psychophysiology, 18*(3), 307-315.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10,* 433-436.

Bull, R. & Shead, G. (1979). Pupil-Dilation, Sex of Stimulus, and Age and Sex of Observer. *Perceptual and Motor Skills, 49*(1), 27-30.

Cai, H. & Wang, J. T. (2006). Overcommunication in Strategic Information Transmission Games. *Games and Economic Behavior, 56*(1), 7-36.

Camerer, C. F., Ho, T.-H., & Chong, J.-K. (2004). A cognitive hierarchy model of games. *Quarterly Journal of Economics, 119*(3), 861-898.

Camerer, C. F., Johnson, E. J., Rymon, T., & Sen, S. (1993). Cognition and Framing in Sequential Bargaining for Gains and Losses. In K. G. Binmore, A. P. Kirman & P. Tani (Eds.), *Frontiers of Game Theory* (pp. 27-47). Cambridge: MIT Press.

Chapman, C. R., Oka, S., Bradshaw, D. H., Jacobson, R. C., & Donaldson, G. W. (1999). Phasic pupil dilation response to noxious stimulation in normal volunteers: Relationship to brain evoked potentials and pain report. *Psychophysiology, 36*(1), 44-52.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The Psychology of Human-Computer Interaction*: L. Erlbaum Associates Inc.

Chen, Y. (2007). Perturbed Communication Games with Honest Senders and Naive Receivers. Arizona State University.

Chen, C.-T., Huang, C.-Y., & Wang, J. T.-y. (2009). A Window of Cognition: Eyetracking the Decision-Making Process in the Graphical Beauty Contest Game.

Conati, C. & Merten, C. (2007). Eye-tracking for user modeling in exploratory learning environments: An empirical evaluation. *Knowledge-Based Systems, 20*(6), 557-574.

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments & Computers, 34*, 613-617.

Costa-Gomes, M., Crawford, V. P., & Broseta, B. (2001). Cognition and Behavior in Normal-form Games: An Experimental Study. *Econometrica, 69*(5), 1193-1235.

Costa-Gomes, M. A. & Crawford, V. P. (2006). Cognition and Behavior in Two-Person Guessing Games: An Experimental Study. *American Economic Review, 96*(5), 1737-1768.

Crawford, V. P. (2008). Look-ups as the Windows of the Strategic Soul: Studying Cognition via Information Search in Game Experiments. In A. Caplin & A. Schotter (Eds.), *The Foundations of Positive and Normative Economics: A Handbook* (Vol. 1, pp. 249-280): Oxford University Press.

Crawford, V. P. & Sobel, J. (1982). Strategic Information Transmission. *Econometrica, 50*(6), 1431-1451.

de Greef, T., Lafeber, H., van Oostendorp, H., & Lindenberg, J. (2009). Eye Movement as Indicators of Mental Workload to Trigger Adaptive Automation *Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience* (pp. 219-228): Springer-Verlag.

Dehais, F., Causse, M., & Pastor, J. (2008). *Embedded eye tracker in a real aircraft: new perspectives on pilot/aircraft interaction monitoring* 3rd International Conference on Research in Air Transportation (ICRAT 2008).

Dionisio, D. P., Granholm, E., Hillix, W. A., & Perrine, W. F. (2001). Differentiation of deception using pupillary responses as an index of cognitive processing. *Psychophysiology, 38*(2), 205-211.

Duchowski, A. T. (2007). *Eye Tracking Methodology: Theory and Practice* (2 ed.). London: Springer-Verlag.

Ekman, I., Poikola, A., Mäkäräinen, M., Takala, T., & Hämäläinen, P. (2008). *Voluntary pupil size change as control in eyes only interaction*. Proceedings of the 2008 symposium on Eye tracking research & applications.

Ekman, I. M., Poikola, A. W., & Mäkäräinen, M. K. (2008). *Invisible eni: using gaze and pupil size to control a game*. CHI '08 extended abstracts on Human factors in computing systems.

Fischbacher, U. (2007). z-Tree: Zurich Toolbox for Readymade Economic Experiments. *Experimental Economics, 10*(2), 171-178.

Frank, M. G. & Ekman, P. (1997). The Ability to Detect Deceit Generalizes Acrosss Different Types of High-Stake Lies. *Journal of Personality and Social Psychology, 72*(6), 1429-1439.

Gabaix, X., Laibson, D., Moloche, G., & Weinberg, S. (2006). Information Acquisition: Experimental Analysis of a Boundedly Rational Model. *American Economic Review, forthcoming*.

Gilzenrat, M., Cohen, J., Rajkowski, J., & Aston-Jones, G. (2003). Pupil dynamics predict changes in task engagement mediated by locus coeruleus. *Society for Neuroscience Abstracts*, No.515.519.

Gneezy, U. (2005). Deception: The Role of Consequences. *American Economic Review, 95*(1), 384-394.

Goldwater, B. C. (1972). Psychological Significance of Pupillary Movements. *Psychological Bulletin, 77*(5), 340-355.

Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica, 57*(2), 357-384.

Heilveil, I. (1976). Deception and Pupil Size. *Journal of Clinical Psychology, 32*(3), 675-676.

Hess, E. H. (1972). Pupillometrics. In N. S. Greenfield & R. A. Sternbach (Eds.), *Handbook of Psychophysiology* (pp. 491-531): Holt, Rinehart & Winston.

Hess, E. H. & Polt, J. M. (1960). Pupil Size as Related to Interest Value of Visual Stimuli. *Science, 132*(3423), 349-350.

Hess, E. H. & Polt, J. M. (1964). Pupil Size in Relation to Mental Activity during Simple Problem-Solving. *Science, 143*(3611), 1190-1192.

Hicks, R. A., Reaney, T., & Hill, L. (1967). Effects of Pupil Size and Facial Angle on Preference for Photographs of a Young Woman. *Perceptual and Motor Skills, 24*(2), 388-&.

Hunton, J. E. & McEwen, R. A. (1997). An Assessment of the Relation Between Analysts' Earnings Forecast Accuracy, Motivational Incentives and Cognitive Information Search Strategy. *Accounting Review, 72*(4), 497-515.

Hurkens, S. & Kartik, N. (2009). Would I Lie To You? On social preferences and lying aversion. *Experimental Economics, 12*(2), 180-192.

Hyona, J., Tommola, J., & Alaja, A.-M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *Quarterly Journal of Experimental Psychology, 48A*, 598-612.

Janisse, M. P. (1973). Pupil Size and Affect - Critical Review of Literature since 1960. *Canadian Psychologist, 14*(4), 311-329.

Janisse, M. P. (1977). *Pupillometry: The psychology of the pupillary response*. Washington, DC: Hemisphere.

Janisse, M. P. & Bradley, M. T. (1980). Deception, Information and the Pupillary Response. *Perceptual and Motor Skills, 50*(3), 748-750.

John, B. (1995). Why GOMS? *interactions, 2*(4), 80-89.

John, B. E. & Kieras, D. E. (1996). The GOMS family of user interface analysis techniques: comparison and contrast. *ACM Trans. Comput.-Hum. Interact., 3*(4), 320-351.

Johnson, E. J., Camerer, C., Sen, S., & Rymon, T. (2002). Detecting Failures of Backward Induction: Monitoring Information Search in Sequential Bargaining. *Journal of Economic Theory, 104*(1), 16-47.

Just, M. A. & Carpenter, P. A. (1993). The intensity dimension of Thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology, 47*, 310-339.

Just, M. A., Carpenter, P. A., Keller, T. A., Eddy, W. F., & Thulborn, K. R. (1996). Brain actication modulated by sentence comprehension. *Science, 274*, 114-116.

Kahneman, D. & Beatty, J. (1966). Pupil Diameter and Load on Memory. *Science, 154*(3756), 1583-1585.

Kahneman, D., Tursky, B., Shapiro, D., & Crider, A. (1969). Pupillary, heart rate, and skin resistance changes during a mental task. *Journal of Experimental Psychology, 79*, 164-167.

Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T.-y., et al. (2009). The Wick in the Candle of Learning: Epistemic Curiosity Activates Reward Circuitry and Enhances Memory. *Psychological Science, 20*, 963-973.

Kartik, N. (2008). Strategic Communication with Lying Costs.

Kartik, N., Ottaviani, M., & Squintani, F. (2007). Credulity, lies, and costly talk. *Journal of Economic Theory, 134*(1), 93-116.

Klingner, J., Kumar, R., & Hanrahan, P. (2008). *Measuring the task-evoked pupillary response with a remote eye tracker.* Proceedings of the 2008 symposium on Eye tracking research and applications, Savannah, Georgia.

Kramer, A. F. (1991). Physiological metrics of mental workload: A review of recent progress. In D. L. Damos (Ed.), *Multiple-task Performance* (1 ed., pp. 279-328): Taylor & Francis.

Lin, T., Imamiya, A., Hu, W., & Omata, M. (2007). Combined User Physical, Physiological and Subjective Measures for Assessing User Cost *Universal Access in Ambient Intelligence Environments* (pp. 304-316): Springer-Verlag.

Lin, T., Imamiya, A., & Mao, X. (2008). Using multiple data sources to get closer insights into user cost and task performance. *Interacting with Computers, 20*(3), 364-374.

Lohse, G. L. & Johnson, E. J. J. (1996). A Comparison of Two Process Tracing Methods for Choice Tasks. *29th Hawaii International Conference on System Sciences (HICSS), 4*, 86.

Lubow, R. E. & Fein, O. (1996). Pupillary size in response to a visual guilty knowledge test: New technique for the detection of deception. *Journal of Experimental Psychology-Applied, 2*(2), 164-177.

Marshall, S. P. (2007). Identifying Cognitive State from Eye Metrics *Aviation, Space, and Environmental Medicine, 78*(5), B165-B175.

Muldner, K., Christopherson, R., Atkinson, R., & Burleson, W. (2009). Investigating the Utility of Eye-Tracking Information on Affect and Reasoning for User Modeling *User Modeling, Adaptation, and Personalization* (pp. 138-149): Springer-Verlag.

Oka, S., Chapman, C. R., & Jacobson, R. C. (2000). Phasic pupil dilation response to noxious stimulation: Effects of conduction distance. sex, and age. *Journal of Psychophysiology, 14*(2), 97-105.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437-442.

Oliveira, F. T. P., Aula, A., & Russell, D. M. (2009). *Discriminating the relevance of web search results with measures of pupil size*. Proceedings of the 27th international conference on Human factors in computing systems.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist, 38*(1), 63-71.

Partala, T. & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International Journal of Human-Computer Studies, 59*(1-2), 185-198.

Pomplun, M. & Sunkara, S. (2003). *Pupil Dilation as an Indicator of Cognitive Workload in Human-Computer Interaction.* Proceedings of the International Conference on Human-Computer Interaction.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372-422.

Recarte, M. A. & Nunes, L. M. (2003). Mental Workload While Driving: Effects on Visual Search, Discrimination, and Decision Making. *Journal of Experimental Psychology / Applied, 9,* 119-137.

Recarte, M. Á., Pérez, E., Conchillo, Á., & Nunes, L. M. (2008). Mental Workload and Visual Impairment: Differences between Pupil, Blink, and Subjective Rating. *The Spanish*

Journal of Psychology 11(2), 374-385.

Sánchez-Pagés, S. & Vorsatz, M. (2007). An Experimental Study of Truth-telling in a Sender-Receiver Game. *Games and Economic Behavior, 61*(1), 86-112.

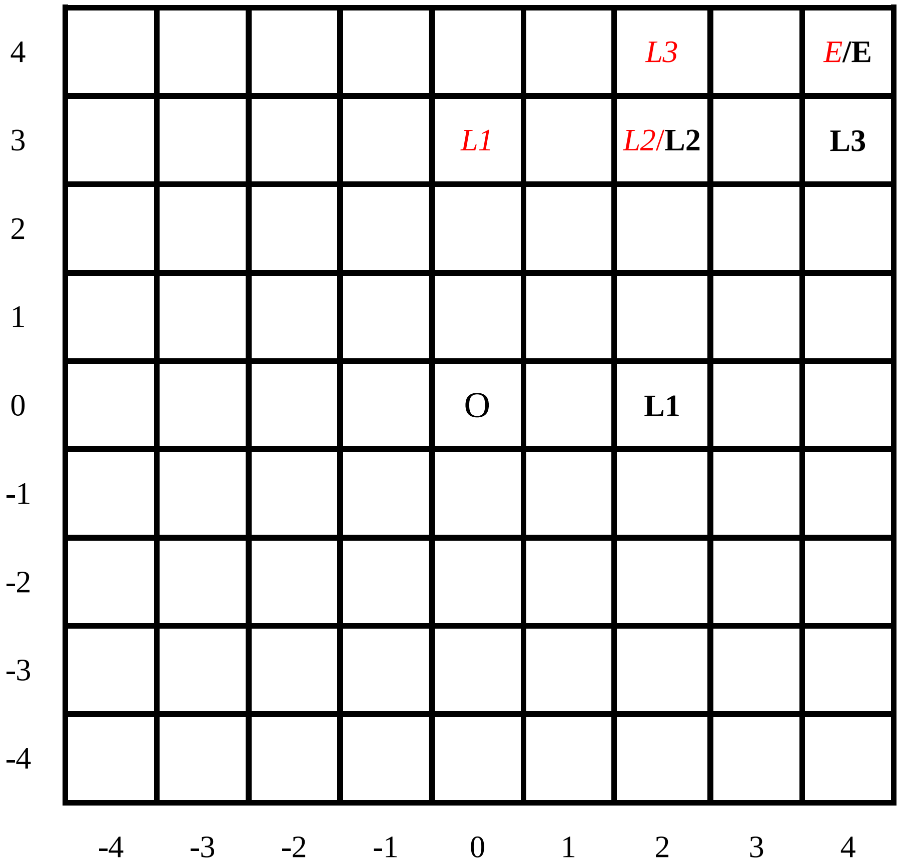
Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preferences. *Nature Neuroscience, 6*(12), 1317-1322.

Siegle, G. J., Steinhauer, S. R., Stenger, V. A., Konecky, R., & Carter, C. S. (2003). Use of concurrent pupil dilation assessment to inform interpretation and analysis of fMRI data. *NeuroImage, 20*(1), 114-124.

Spence, S. A., Farrow, T. F. D., Herford, A. E., Wilkinson, I. D., Zheng, Y., & Woodruff, P. W. R. (2001). Behavioural and Functional Anatomical Correlates of Deception in Humans. *NeuroReport, 12*(13), 2849-2853.

Wang, J. T.-y., Spezio, M., & Camerer, C. F. (2009). Pinocchio's Pupil: Using Eyetracking and Pupil Dilation To Understand Truth-telling and Deception in Sender-Receiver Games. *American Economic Review, forthcoming*.

Figures and Tables



4							*L3*		*E*/**E**
3					*L1*		*L2*/**L2**		**L3**
2									
1									
0					O		**L1**		
-1									
-2									
-3									
-4									
	-4	-3	-2	-1	0	1	2	3	4

Figure 1: Equilibrium and Level-k Predictions of a 9x9 Spatial Beauty Contest Game with Goals (2, 0) and (0, 3).

Appendix: Conducting an Eye-Tracking Study using Eyelink II or Eyelink Remote

In this section, we describe how an eye-tracking is conducted using specific eye-trackers, namely the Eyelink Remote support-free eye-tracking system (SR Research, Osgoode, Ontario) and the mobile Eyelink II head-mounted eye-tracking system (SR Research, Osgoode, Ontario).

The Eyelink Remote is a pair of tiny cameras mounted on a mobile stand placed in front of the computer screen facing toward the subjects' eyes. One camera tracks a target sticker attached to subjects' forehead and captures head movements, while the other tracks one eye and captures eye movements as well as pupil size. Subjects can move their heads and a period of calibration adjusts for head movement to infer accurately where the subject is looking.

The mobile Eyelink II is an older model consists of a pair of tiny cameras mounted on a lightweight rack facing toward the subjects' eyes, and supported by comfortable head straps. Subjects can move their heads and a period of calibration adjusts for head movement to infer accurately where the subject is looking. In addition, the mobile Eyelink II includes an option to install a "scene camera" which captures what subjects see. A "video overlay" package allows one to overlay eye-tracking position onto the video later. This is useful for experiments that are not suitable to conduct on a computer screen, such as face-to-face bargaining, facial lie-detection, or grocery shopping.

Both models are capable of tracking at high speeds (either 250 or 500Hz), allowing for sophisticated gaze-contingent experiments, and share the same data analysis and experimental software. However, Eyelink Remote records under a more natural setting since no device is attached to the subject, while wearing the mobile Eyelink II for long periods of time could be painful due to the weight of the head mount. On the other hand, the mobile Eyelink II has the advantage of not being restricted to computer screen display (with its scene camera and video overlay option).

Computerized experiments (run in Windows XP or Vista) can be created using the Experimental Builder provided by the manufacturer, or programmed by the experimenter in Matlab (Mathworks, Inc., Natick, MA) using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997), which includes the Eyelink Toolbox (Cornelissen Peters, & Palmer, 2002).[9] Programming in other software is also possible since drivers are provided on the manufacturer's supporting forum.[10] These programs would be run on the "display computer" to display the content subjects would see on their computer screen. The "host computer" is run under DOS mode to control the eye-tracker. Before starting a session, the demo program for gaze-contingent experiments, or "GC window", should be performed when the eye-tracker and the display computer are both up running. This demo program calls the appropriate routines to establish the link between the eye-tracker and the display computer. For gaze-contingent experiments, it is suggested that the experiment should be displayed on large high-refreshing (85-100Hz) CRT monitors instead of LCD panels.[11]

When subjects come in, they need to be "connected" to the eye-tracker. For Eyelink II, this means wearing the helmet, tightening the straps, and possibly tying up long hair. For Eyelink Remote, this only requires placing the target sticker on the subject's forehead (since it is

---

[9] See http://psychtoolbox.org. Note that version 3 of the Psychophyics Toolbox (PTB-3) incorporated the Eyelink Toolbox in its final release, while version 2.54 (PTB-2) requires one to install the Eyelink Toolbox separately.

[10] One problem with most existing software is they are mainly designed for experiments on individual decision making, and hence, lack the ability to communicate between eye-trackers. This limits the use of eye-tracking in experiments with strategic interaction. Moreover, random matching could be difficult to conduct. This is a reason why Chen et al (2009) focus solely on initial response (without feedback). Wang et al (2009) and Knoepfle et al (2009) both employed zTree (Zurich Toolbox for Readymade Economic Experiments) developed by Fischbacher (2007) to perform the experiments for all untracked subjects and manually transferred the data between the Psychophysics Toolbox and zTree.

[11] Due to physical limitations of liquid crystals, most LCD panels have a refresh rate of 60-75Hz. This would cause visual delay in gaze-contingent experiments.

remote), and adjusting camera locations. Then, nine-point calibrations (described earlier) and validations (re-fixating on the same baseline location for the eye-tracker to confirm) are performed. Accuracy in the validations should be better than 0.5° of visual angle. Otherwise, re-calibration or further adjustment may be needed.[12] Experimental subjects should not wear contact lenses (unless the lenses would not move during the entire experiment), since contact lenses created an additional cornea reflection which would shift whenever the lenses shift.[13] To ensure accuracy during the experiment, new nine-point calibrations and validations are typically performed prior to the start of each experiment, and drift corrections (Fixating on a black dot in the center of the screen for the eye-tracker to realign calibration) are performed before every trial.

After the experiment is conducted, the eye-tracking data can be analyzed using the Eyelink Data Viewer (SR Research, Hamilton, Ontario).[14] In discriminating fixations, one can typically set saccade velocity, acceleration, and motion thresholds to 30°/sec, 9500°/sec^2, and 0.15°, respectively. Regions of interest (ROIs), or the boxes subject look up, can be drawn on each task image using the drawing functions within the Data Viewer. Measures of gaze included lookup counts, or fixation number (the total number of fixations within an ROI), and fractional lookup time or dwell time (the time during a given round spent fixating a given ROI divided by the total time between image onset and response). Only those fixations beginning between 50ms following the onset of a task image and offset of the task image should be considered for analysis.

---

[12] Adjustments include changing the angle or moving the location of the camera, or simply performing the calibration again. However, in some cases, it may be impossible to calibrate a particular subject, and the experimenter may need to replace this particular subject. Issues that may occur include subjects having a lazy eye (though switching to only tracking the normal eye may help), and subjects having their eyes only "half opened", either due to personal habit or sleepiness (though a good night sleep or consciously keeping one's eyes wide open may help).

[13] On the other hand, glasses are fine as long as one can avoid the additional cornea reflection created by the glass. This is feasible since glasses do not directly contact the corneal.

[14] Constantly upgraded and the demo version freely available at the support forum of http://www.eyelinkinfo.com.

The Data Viewer is capable of importing screenshots (saved by the experimental software during the experiment) and overlaying fixations and saccades on them. These fixation maps and saccade maps summarize the lookup sequences and provide powerful visualization of the entire decision making process, and can be saved as figures. Furthermore, the Data Viewer can export video clips of animated replay of the entire sequence of eye movements. Finally, the Data Viewer can summarize the data and create different "reports" in plain text format. Examples of these reports include the Sample Report (raw data), the Trial Report (summary statistics of each trial), the Interest Area Report (summary results of each ROI), the Fixation Report (information of each fixation), and the Saccade Report (information of each saccade). These reports can be imported into statistical packages such as STATA or MATLAB for further analysis.