

Combining Pupil Diameter Change and Galvanic Skin Response with the Moral Directed Lie Comparison Question Technique

An Evaluation of Its Effectiveness in Lie Detection

Ouyang Tai-Ju, Huang, Yi-Siou, Kuo, Kuan-Ting, et al.

Criminal Investigation Bureau, National Police Agency, Ministry of the Interior

Contact: Mobile 0912992388; email: allen22@cib.npa.gov.tw

Abstract

In modern criminal investigation practice, front-line investigators often face the dual challenges of analyzing large volumes of information and meeting case-closure deadlines, and they frequently find themselves working alone on their cases. Confronted with the broad scope of investigative work, they often feel willing but unable to keep up. For the officer in charge of a case, the most pressing need is to confirm whether the direction of the investigation is correct, so as to avoid the futile investment of resources. The truth of a case, however, often exists only in the mind of the suspect; if it could be glimpsed through polygraph testing, this would facilitate precise criminal investigation. Although polygraph techniques can help clarify the truth and guide the direction of an investigation, current techniques—because of their cumbersome procedures, time-consuming pre-test interviews, and high professional threshold—are usually confined to back-end forensic examination and can rarely be deployed in the golden early stage of an investigation. In addition, current contact-type sensors are cumbersome to wear and susceptible to interference from physical countermeasures, further increasing the difficulty of practical application and the complexity of chart interpretation. Therefore, to help investigators quickly clarify cases and reduce the burden on front-line personnel, it is necessary to develop a polygraph technique with “rapid screening” characteristics that can be transformed from a back-end forensic tool into a front-end investigative aid.

To overcome this practical dilemma, this study explores a technical pathway with “rapid screening” potential, adopting the faster and more easily standardized “Moral Directed Lie Comparison Question” combined with the “Utah Zone Comparison Technique (Utah ZCT)”, and using a non-contact pupillary physiological channel and galvanic skin response for comparative analysis. This simplifies the psychological guidance process in the current pre-test interview and improves the standardization of test administration, while the introduction of the pupillary channel takes advantage of its non-invasiveness and stable accuracy to reduce the physical restraint and psychological interference experienced by examinees while maintaining a comparable level of accuracy.

The findings are as follows. First, based on the experimental data, the Moral Directed Lie Comparison Question combined with the Utah Zone Comparison

Technique (Utah ZCT) produced significant differences between liars and truth-tellers in the two physiological channels of pupil diameter change and galvanic skin response (d values of 0.55 and 0.75, respectively). Second, the specificity of pupil diameter change measurement reached 0.89, effectively excluding innocent examinees. Third, an analysis model integrating the pupillary and electrodermal physiological indicators outperformed either single indicator in diagnostic validity.

This study confirms that, by integrating the directed lie technique with the pupillary channel and using two channels to assess whether an examinee is lying, the accuracy of the MSRD system's automated scoring reached 70%, while the accuracy of manual scoring reached 80.56%. This demonstrates that the directed lie technique not only lowers the threshold for applying polygraph techniques but also maintains a considerable degree of accuracy while shortening test time. This model not only helps ease the caseload of investigative agencies, but also transforms polygraph testing from back-end forensic examination into a front-end investigative screening tool, helping first-line personnel quickly clarify cases while ensuring scientific accuracy and achieving the precise allocation of judicial resources.

Keywords: cognitive load, interview EDA technique, deception indicators

1. Introduction

The comparison questions used in the Comparison Question Technique include the Probable Lie Comparison Question (hereinafter PLC) and the Directed Lie Comparison Question (hereinafter DLC). Regarding the former, Shurany and Gordon (2021) cited Charles Honts's 2015 criticism that the PLC requires the polygraph examiner to possess considerable psychological sensitivity, maturity, and skill in order to obtain accurate results[1]. Regarding the latter, Honts (1994) argued that an important advantage of the DLC is its higher degree of standardization, noting that with the DLC every examinee can receive the same directed lie comparison questions. Because the DLC interview procedure does not rely on the inference that an innocent examinee will probably lie on the comparison questions, the DLC is regarded as simpler and more direct[2]. Krapohl et al. (2023), in "Terminology reference for the science of psychophysiological detection of deception", wrote of the DLC: "Although there are concerns that the DLC may be susceptible to countermeasures, its real advantages are that the DLC is easier to standardize than the PLC, is less intrusive, and its validity depends less on the examiner's level of skill." [3]. However, Iacono (2024) contended that the DLC has still not truly achieved standardization, one reason being that different examiners may use different DLCs, and examinees choose for themselves the lies they will tell during the DLC interview procedure; the significance of these lies to them is unknown and evidently varies from examinee to examinee[4]. Rashid et al. (2024) proposed the Moral Directed Lie Comparison Question (hereinafter MDLC), replacing questions based on life experience with questions involving moral scenarios, aiming to establish a more consistent test and to prompt examinees toward more rational thinking. Rashid et al. (2024) considered that although the MDLC showed some potential in preliminary trials, its validity relative to the DLC awaits further evaluation in subsequent research[5]. This paper adopts the question format of the Utah technique, replacing its comparison questions with MDLCs, to observe the differences in galvanic skin response and pupil diameter change between liars and truth-tellers under this question format.

2. Literature Review

According to the “Terminology reference for the science of psychophysiological detection of deception”, the polygraph scholar Backster introduced the term psychological set, regarding it as the core concept of the comparison question technique and linking it to the emotion of fear. According to Backster’s hypothesis, examinees pay more attention to the types of questions that pose the greatest threat to their interests. Examinees who lie to the relevant questions perceive those questions as more threatening than the others, and therefore attend to them more and exhibit greater physiological arousal. Likewise, innocent examinees find the probable-lie comparison questions more threatening, devote more attention to them, and consequently produce greater arousal responses. Nelson (2015) discussed the psychological foundations of polygraph testing, arguing that the concept of psychological set is unfalsifiable and unmeasurable [6], and that it should be replaced by the concept of “differential salience” proposed by Senter et al. (2010), which explains the physiological responses evoked by polygraph question stimuli in terms of emotion, cognition, and conditioned learning [7]. Nelson (2015) began his account with Pavlov’s (1927) conditioned behavior, supplemented by subsequent learning theory, including the concepts of stimulus sensitization and habituation [6]. From this perspective, an examinee’s response to polygraph question stimuli can be viewed as a form of conditioned learning: a complete and effective pre-test interview allows a truthful examinee to habituate to the test questions; conversely, for a liar, these questions become sensitizing, acting as conditioned stimuli that evoke stronger responses. In addition, Nelson (2015) considered that cognitive-behavioral theory, encompassing cognition, emotion, behavior, and experiential learning, can serve as a hypothesis to explain the various known phenomena of polygraph testing (Kahn, Nelson & Handler, 2009) [8]. From a more general cognitive-behavioral perspective, telling the truth involves relatively simple cognitive and emotional load, whereas lying is more complex. Nelson (2015) further argued that, according to the cognitive-behavioral and differential-salience models, for a set of repeatedly presented polygraph test stimuli, physiological responses will show different loads on different types of test questions depending on whether the examinee is truthful or deceptive regarding the test issue. These responses can be understood as arising from the examinee’s cognition, memory, emotion, and past conditioning experiences with respect to the test question stimuli. Differences in relative responses across stimuli can be compared against statistical reference distributions and assessed for statistical significance, quantifying the degree of uncertainty in reaching a classification of deception or truthfulness.

Since the physiological responses obtained from administering a series of test questions can be used to compute statistical differences between liars and truth-tellers, the choice of physiological channel is very important. Among the physiological channels currently used for deception detection, those recognized by the American Polygraph Association are respiration, electrodermal response, and cardiovascular recording; pupil diameter change has not yet been included, but a considerable body of literature indicates that pupil diameter change has potential for identifying deception. Kahneman and Beatty (1966) initiated research on mental effort and pupil diameter change. They placed subjects in high-incentive (10 cents) versus low-incentive (2 cents) and high-difficulty versus low-difficulty

experimental conditions, and found that incentive level did not affect pupil diameter change, whereas pupil diameter dilated significantly as the difficulty of tasks involving mental effort increased [9]. Lubow and Fein (1996) conducted an experiment resembling the Guilty Knowledge Test on 28 Tel Aviv University students, collecting pupil diameter changes. Subjects who simulated the crime of forging negotiable securities were tested with question sets each containing six slides—five control slides and one target slide—for a total of four sets (signature, ID badge, amount, and face). The pupil diameter after each slide stimulus was measured. For threshold scoring of concealed information, Lykken's 1959 scoring method was applied to the magnitude of pupil diameter change: the largest change received 2 points, the second largest 1 point, and the rest 0 points, and each subject's total score was computed (maximum 8 points per question set, minimum 0). The results showed that among the guilty group, the difference in mean pupil diameter between target and control slides was significantly greater for the deceptive group than for the innocent group; using 4 points as the decision threshold, 70% of the guilty and 100% of the innocent were correctly identified. Lubow and Fein (1996) then added galvanic skin response sensors to the subjects. In a second mock-crime experiment, pupil diameter change correctly identified 50% of the guilty and 100% of the innocent, while galvanic skin response identified 55% of the guilty and 93% of the innocent, but Lubow and Fein (1996) did not analyze the accuracy of combining galvanic skin response with pupil diameter change [10]. Webb, Honts, et al. (2009) at the University of Utah used the same comparison question format as the Utah technique to evaluate the feasibility of using pupil diameter as a deception indicator in the context of a comparison question polygraph test, to examine which combination of pupil diameter with the commonly used respiration, electrodermal, and cardiovascular channels would achieve the best results, and to discuss the possibility of pupil diameter replacing one of the polygraph channels. The study used a mock theft paradigm and collected data from 24 subjects (12 innocent and 12 guilty). The results showed that for innocent subjects, pupil dilation was significantly greater for comparison questions than for relevant questions, while guilty subjects showed no significant difference between the two question types. Webb, Honts, et al. (2009) considered this consistent with the theoretical basis of the CQT. Because lying requires more cognitive load than answering truthfully, and pupil diameter reflects changes in cognitive load, innocent subjects lie only on the comparison questions, so pupil dilation should be more pronounced on comparison questions than on relevant questions. On the other hand, guilty subjects showed no significant difference in pupil dilation between comparison and relevant questions—presumably because they had to respond deceptively to both question types, and their pupil responses may simply reflect this. Webb, Honts, et al. (2009) also compared regression models across the physiological channels: combining pupil diameter with the electrodermal and respiration channels produced the highest R^2 value ($R^2=0.49$), which was 0.09 higher than the R^2 value ($R^2=0.40$) of the combination of pupil diameter with the electrodermal and cardiovascular channels. Webb, Honts, et al. (2009) concluded from these results that pupil diameter has the potential to replace the cardiovascular channel[11]. In recent research on pupil diameter change, the scholars John Kircher, Doug Hacker, and colleagues have explored whether eye movements and pupil responses can serve as indicators for detecting deception in computerized tests (Cook et al., 2024). Cook et al. (2024) described the theoretical framework of John Kircher, Doug Hacker, and

colleagues as built on four core assumptions: lying carries a high cognitive load, lying elicits emotional arousal, lying requires self-regulation, and lying depends on self-motivation[12]. A large body of research consistently indicates that deception is more cognitively demanding than truthful statements (Johnson Jr et al., 2005; Kircher, 1981; Lubow & Fein, 1996; Vrij, 2008; Walczyk et al., 2009; Walczyk et al., 2003) [13] [14] [15] [16] [17]. During lying, the individual must comprehend the question, inhibit the true answer, construct a plausible lie, maintain narrative consistency, and simultaneously attend to the consequences of the lie being detected. These mental processes significantly increase cognitive load and manifest as pupil dilation (Ahern & Beatty, 1979; Beatty & Wagoner, 1978; Colman & Paivio, 1970; Hess & Polt, 1964; Kahneman & Beatty, 1966) [9] [18] [19] [20] [21]; Dionisio et al. (2001) found that individuals show more pronounced pupil dilation when fabricating lies than when answering truthfully, which the researchers attributed mainly to the higher cognitive processing load required by lying[22]. Subsequent research also supports the association between pupil response and deception: in situations of deception or information concealment, pupil diameter is generally larger than when answering honestly (Bradley & Janisse, 1981; Heilveil, 1976; Webb, Hacker, et al., 2009; Webb, Honts, et al., 2009) [23] [24] [25] [11]. Notably, pupil dilation may be influenced by both cognitive load and emotional arousal, so the pupil changes observed in deception contexts may reflect the individual simultaneously bearing higher cognitive processing demands and emotional stress while concealing the deception (Cook et al., 2024) [12]; effective deception also relies on self-regulation, including continuous monitoring and control of thoughts, emotions, and behavior (Zimmerman, 2008) [26]. Deceivers must suppress natural emotional reactions, track inconsistencies in their fabricated statements, and continually adjust their behavior to maintain credibility. These ongoing self-monitoring behaviors increase cognitive load, leading to further pupil dilation; sustaining deception requires strong self-motivation. Liars must believe they can deceive successfully, expect to be believed, and maintain prolonged focus and effort—a process that includes monitoring responses, inhibiting truthful reactions, and managing emotions, all of which increase cognitive load (Zimmerman & Moylan, 2009) [27]. Consequently, liars typically show longer response times, more pronounced pupil dilation, and reading behavior patterns that differ from those of truth-tellers.

Another important physiological response channel is the galvanic skin response (Galvanic Skin Response, also known as Electrodermal Activity, hereinafter EDA). The galvanic skin response is a method of measuring skin conductance (Boucsein et al., 2012) [28], and is a powerful technique that can be measured in real time and is linearly related to the level of arousal (Rahma et al., 2022). EDA measures changes in skin conductance, which depend on the amount of sweat secreted by the eccrine sweat glands in the subcutaneous layers of the palms and soles. Sweat secretion in the palms and soles arises mainly from neural activity related to affective and cognitive states, including psychological or emotional sweating (Rahma et al., 2022) . The EDA signal consists of a tonic component and a phasic component. The tonic component relates to the slowly varying background characteristics of the signal and is also called the Skin Conductance Level (SCL). In contrast, the phasic component is the rapidly changing phenomenon in the signal, also called the Skin Conductance Response (SCR), because it appears only within specific time windows; an SCR occurs when the sympathetic nervous

system responds to certain stimuli. The SCR can therefore be linked to specific stimuli and used to gauge cognition, emotional arousal, or stress levels (Rahma et al., 2022). However, the individual components of EDA cannot be extracted directly from the signal. The signal measured by an EDA sensor is presented as raw Skin Conductivity (SC), usually in microsiemens (Rahma et al., 2022) [29]. The comparison of the electrodermal response with the other two physiological channels—respiration and cardiovascular activity—has long been debated in the polygraph field. Widacki (2015) concluded from many laboratory studies that the electrodermal response is the most valuable indicator, followed by the cardiovascular response, with respiration being the weakest. Meijer et al. (2014) conducted a meta-analysis comparing the validity of electrodermal response, heart rate, respiration line length (RLL), and the P300 brain wave in laboratory studies across two settings—personal-item tests (e.g., tests of the examinee's surname, given name, or date of birth) and mock crimes. The results showed that the P300 outperformed the electrodermal response only in personal-item tests, not in mock crimes; when only electrodermal response, heart rate, and respiration line length were compared, the electrodermal response still showed the superior effect size[30]. Nelson (2017) synthesized numerous studies (Ansley & Krapohl, 1999; Honts, Handler, Shaw & Gougler, 2015; Harris, Horner & McQuarrie, 2000; Kircher, Kristjansson, Gardner & Webb, 2005; Kircher and Raskin, 1988; Krapohl & McManus, 1999; Nelson, Krapohl & Handler; Podlesny & Truslow, 1993; Raskin, Kircher, Honts & Horowitz, 1998) and found that the electrodermal response carries a higher weight in scoring than the other physiological channels; accordingly, in the Empirical Scoring System (ESS) he developed, all electrodermal integer scores are doubled. Reflected in the scoring of a single chart, EDA scores are -2, 0, +2, while the other physiological channels score -1, 0, +1. All of this shows that the electrodermal response occupies a very important position in polygraph research and practice[31].

Summarizing the studies above, the electrodermal response and pupil diameter change both perform at a considerable level in deception detection, but whether comparing them individually or combining the two can achieve higher accuracy awaits further research. In addition, all polygraph tests use a standardized Question Format; when the comparison question technique is adopted, either probable lie or directed lie comparison questions may be used. Rashid et al. (2024), within the framework of the directed lie comparison question, proposed the Moral Directed Lie Comparison Question (hereinafter MDLC); the relevant context is explained below[5]:

According to the U.S. National Research Council (2003), the technique commonly used by polygraph examiners is the Comparison Question Technique (CQT). The main concept of the CQT is to place the examinee's responses within a series of psychologically meaningful questions for comparison, thereby determining whether the physiological responses to the relevant questions are significantly abnormal. The comparison questions used in the CQT fall into two categories: the PLC and the DLC. Menges (2004) synthesized the empirical data accumulated since 1987 and found that polygraph examinations using the DLC have good accuracy. Compared with the PLC, the DLC has several advantages. It is less intrusive on privacy, easy to standardize, and avoids the difficulties of developing PLCs that commonly arise in repeated testing. For examiners who must conduct large numbers of repeated tests, these characteristics have considerable practical significance. Menges (2004) stated that the earliest discussion of the DLC came from an unpublished document written by

former examiners of the U.S. Army Military Intelligence (MI) polygraph program. In June 1982, at the Federal Interagency Polygraph Seminar (FIPS) held in Quantico, Virginia, Louis Fuse and his colleagues explained the background, rationale, and manner in which examiners had developed and used the DLC in practice. Fuse's unpublished manuscript "Directed Lie Control Testing Technique" is regarded as one of the seminal introductions of the DLC technique in the polygraph literature[32].

Regarding accuracy, taking the Utah technique paired with the directed-lie question format as an example, Honts and Raskin (1988) analyzed the accuracy of Utah DLC tests of 25 criminal suspects (12 deceptive and 13 truthful), using confessions, evidence, confessions of other suspects, or withdrawal of complaints as the criterion, obtaining an accuracy of .958 with an inconclusive rate of 0.077[33]. Horowitz et al. (1997) conducted a laboratory study using the Utah DLC test, with 15 subjects assigned to lie and 15 assigned to be truthful. The results yielded an accuracy of .856 and an inconclusive rate of 0.067[34], demonstrating that the combination of the Utah technique with directed lie comparison questions has very good validity in both laboratory settings and real criminal cases. However, although the directed lie question has the advantages of lower privacy intrusion and easy standardization, as well as respectable validity, as noted in the introduction, Iacono (2024) maintained that the DLC has still not truly been standardized: different examiners may adopt different DLCs, and examinees choose their own lies during the DLC pre-test interview; the significance of these lies to the examinee is unknown and clearly varies across examinees[4]. Rashid et al. (2024) tested 74 subjects with the DLC (37 in the deceptive group, 37 in the truthful group) and found significant differences in pupil diameter change between the two groups. However, Rashid et al. (2024) reported from the feedback of most subjects that they were confused by the wording and design intent of the DLC items, and occasionally answered contrary to the test instructions—especially because the items usually involve personal experience, and individuals differ in whether they have had relevant lying experiences—which the authors considered a possible obstacle to standardization. Therefore, Rashid et al. (2024) proposed the MDLC, inspired by dual-process models, which hold that situations involving moral dilemmas simultaneously evoke two distinct but competing mental processes, commonly referred to as System 1 and System 2 (Stanovich, 1999) [36]. System 1 is associated with fast, automatic, emotionally charged processing, whereas System 2 is characterized by more conscious, deliberate, and controlled reasoning. Take Kvaran et al. (2013)'s experiment as an example: one of the moral dilemma questions they used reads as follows[37]:

You are driving a runaway trolley that is rapidly approaching a fork in the track. On the track to the left are five railway workers; on the track to the right is one railway worker. If you do nothing, the trolley will head left, killing the five workers. The only way to avoid the deaths of these five people is to press the switch on your dashboard, diverting the trolley to the right-hand track—but doing so will kill the single worker on that track.

Participants primed for rational analysis were asked by the experimenter to solve a mathematics problem before answering the moral dilemma, while unprimed participants were not. The results showed that those primed for rational analysis tended to choose to press the switch on the dashboard, killing the one worker on the right-hand track. This phenomenon can be explained by Kahneman

(2013)'s dual-system theory: the preceding mathematical task successfully primed System 2, which is responsible for logical computation, enabling the individual to suppress emotional intuition in the subsequent moral choice and lean toward utilitarian, rational weighing. However, the operation of System 2 requires substantial cognitive resources; when an individual is under high cognitive load, the resources available for deliberate reasoning may be limited, increasing the likelihood of reliance on intuitive judgment. Given that moral judgment involves value conflict and cognitive trade-offs[35], Rashid et al. (2024) sought to propose the Moral Directed Lie Comparison Question—retaining the core structure in which the truth is known and the subject is instructed to lie—to replace traditional comparison questions based on life experience, which easily provoke unintended emotional reactions; an example follows[35]:

Do you think that, morally speaking, it is wrong to lie to frame someone?

Because such questions involve moral norms and value beliefs generally accepted by most individuals, when examinees are asked to answer contrary to the question, they must first inhibit the natural response consistent with their true beliefs and continuously maintain and monitor the false response required by the experiment. This process involves response inhibition, working-memory maintenance, and executive control, and may therefore increase the individual's cognitive load. Compared with traditional comparison questions built on personal life experience, the MDLC uses standardized moral-conflict stimuli to reduce the emotional interference caused by individual differences in experience and to improve the consistency and controllability of the test stimuli.

3. Methods

This paper investigates whether pupil diameter change and galvanic skin response, combined with the Moral Directed Lie Comparison Question, can effectively distinguish deceptive from truthful examinees. The research framework and hypotheses, participants, instruments and environment, procedure, and data processing and statistical methods are described in turn below:

(1) Research Framework and Hypotheses

In this study, a lying scenario was simulated in the laboratory. Subjects were divided into an experimental group and a control group. The experimental group was asked to fabricate a life event from a specific recent period (e.g., the weekend), while the control group was asked to describe a real and memorable event from a specific recent period (e.g., the weekend). Both groups wrote the event on the experiment record form, from which the researcher prepared the relevant questions using the Utah comparison question technique format. An open-ended question asked the subject to describe the event in detail to deepen the impression. Based on the record form, the researcher explained in order the Moral Directed Lie Comparison Questions, the irrelevant questions, the introduction question (the Introduction Question in the Utah ZCT, semantically similar to the symptomatic question in the ZCT), and the sacrifice relevant question. After confirming that the subject understood the meaning of the questions, the acquaintance test and the main test (in-test) were administered.

During testing, an eye tracker and a galvanic skin response recorder simultaneously collected physiological signals for subsequent analysis and determination of results. This paper makes the following hypotheses regarding the lie-detection effectiveness of pupil diameter change and galvanic skin response:

1. Using a question format with Moral Directed Lie Comparison Questions, the deceptive and truthful groups differ significantly in the relevant-minus-comparison difference in pupil diameter change and galvanic skin response.
2. The unweighted accuracy of lie detection combining pupil diameter change and galvanic skin response can reach 80% or higher.
3. Combining pupil diameter change with galvanic skin response, compared with using galvanic skin response alone, reduces the inconclusive rate and increases unweighted accuracy.

(2) Participants

This study recruited 81 volunteer students from Central Police University. Excluding one subject whose file was overwritten due to an instrument operation error, the remaining 80 completed the test: 60 men and 20 women,

aged approximately 18–57 years (mean 24.89). Pre-test screening showed no stroke or cognition-affecting illness and no alcohol consumption before testing; average pre-test sleep duration was 8.32 hours. Four subjects reported abnormal physical conditions (colds), but none reported fatigue or inability to concentrate. All subjects had received at least 12 years of basic national education. This study was approved by the Research Ethics Committee of National Changhua University of Education (protocol number: NCUEREC-114-004).

(3) Instruments and Environment

The test instruments were the MSRD (Multi-Spectral Reality Detector, hereinafter MSRD) analysis software designed by the Singapore instrument company AISeer and the galvanic skin response recorder produced by the same company, integrated with a Tobii bar-type eye tracker for recording pupil



Figure 1. Schematic of the MSRD software interface, galvanic skin response recorder, and bar-type eye tracker

diameter changes. Early pupil diameter measurement had to be conducted in low-light or fully dark environments to reduce interference from ambient light on pupil image contrast. In recent years, however, improvements in eye-tracker technology have made it possible to measure pupil diameter accurately under ordinary lighting. The **eye tracker model currently used is the** Tobii Pro Eye Tracker 5L, with a sampling rate of 120 Hz, capable of detecting pupil diameter changes in natural light.

Testing took place in the polygraph room of the Department of Criminal

Investigation at Central Police University. The room is soundproof, free from undue external interference, and has stable air conditioning and indoor lighting, so subjects were tested under stable and suitable temperature and light. As shown in Figure 1, the subject faced a screen; questions were asked by voice from a computer running the software interface provided by AISeer, and the subject answered by operating the mouse. The collected physiological signals were uploaded over the network to AISeer for remote analysis and generation of physiological charts.

(4) Procedure

The researcher first recruited subjects on campus. At the appointed time the subject reported to the polygraph room, where the researcher explained the consent form, the testing environment, and the instruments. After consent was obtained, basic information and physical condition (sleep, medication, illness, etc.) were surveyed, followed by the acquaintance test. After confirming that the subject could correctly answer the questions and was familiar with operating the instruments, the subject drew lots to determine assignment to the deceptive or truthful group. The deceptive-group lot read: “What I state will be fictitious”; the truthful-group lot read: “What I state will be truthful”. After drawing, the subject did not put the lot back into the bag and did not let the researcher know which lot was drawn. The researcher then asked the subject to write, in one sentence on the experiment record form, the most memorable event of the most recent weekend, including the time, place, and people involved. After writing, the subject handed the record form to the researcher, who used it to prepare the relevant questions and, once they were prepared, read each question to the subject for confirmation.

After confirming that the subject understood the relevant questions, the researcher explained the moral comparison questions as follows:

Next I will ask you some questions that require you to think before answering. I ask this because when people handle questions that require thought, the mental effort consumes energy and produces physiological responses. I need to record these normal physiological responses so that they can be compared with your answers to the life-event questions.

After this explanation, the answering method for each question was explained in turn, as follows:

For the first question in this part, I will ask: “Do you think that, morally speaking, it is wrong to lie to frame someone?” “Normally, lying to frame someone is wrong, but here I am asking you to answer in a way that violates your moral views—so what should you answer?”

When the subject could correctly answer “No”, the explanation proceeded to the next moral comparison question, as follows:

“Do you think that, morally speaking, it is acceptable to deliberately lie to others?”

“Normally, one may not deliberately lie to others, but if you are asked to answer in a way that violates your moral views, what should you answer?” The subject answers “Yes”.

“Do you think that, morally speaking, lying to evade responsibility is impermissible?”

“Normally, lying is impermissible, but if you are asked to answer in a way that violates your moral views, what should you answer?” The subject answers “No”.

Once the subject could answer all the moral comparison questions correctly, the irrelevant questions, the symptomatic question, and the sacrifice relevant question were explained, after which the main test began. The main test collected three physiological charts in total. After collection, the data were uploaded on the spot to AISeer’s cloud service and the analysis results returned; the subject was informed and the lot was retrieved, after which the chart records were statistically analyzed. The procedure is shown in Figure 2 below:

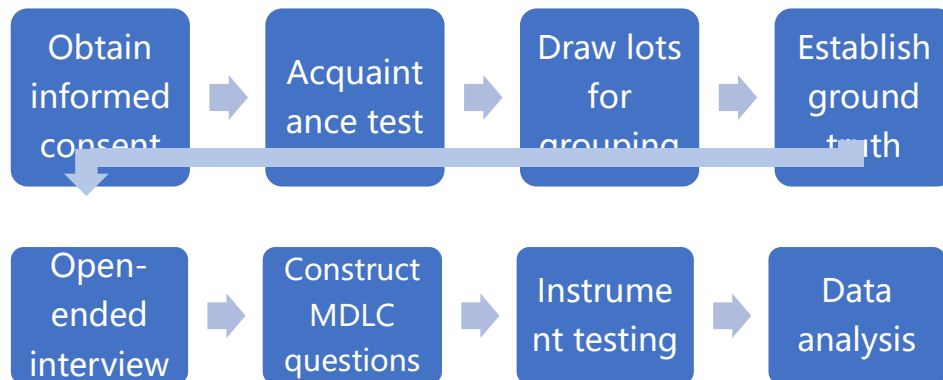


Figure 2. Research flow chart

(5) Data Processing and Statistical Methods

1. Differences in pupil diameter change and galvanic skin response within the response window, and statistical tests

The physiological responses of pupil diameter change and galvanic skin response over the 0–15 second window for the 9 relevant questions and 9 comparison questions across the 3 tests were summed and averaged, and plotted as two charts—one representing the pupil diameter changes of that test and the other the galvanic skin response changes. As shown, each chart has three response curves: one for the physiological response to the relevant questions over 0–15 seconds, one for the comparison questions, and one for the irrelevant questions, as in Figures 3 and 4 below.

Average Pupil Response Post Question Stimuli

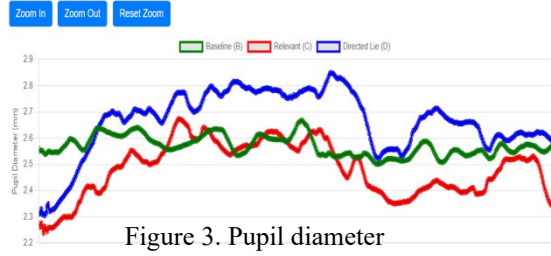


Figure 3. Pupil diameter

change curves

Average GSR Response Post Question Stimuli

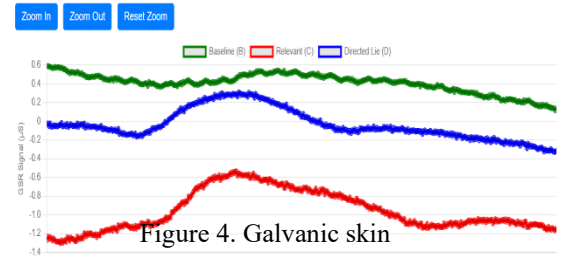
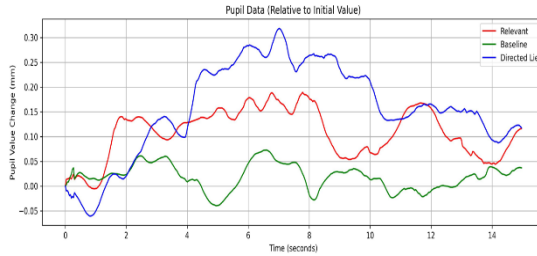


Figure 4. Galvanic skin

response curves

The physiological responses of the three tests for pupil diameter change and galvanic skin response were first superimposed and averaged, yielding one average physiological response curve each for the relevant, comparison, and irrelevant questions, as shown below.



change curves

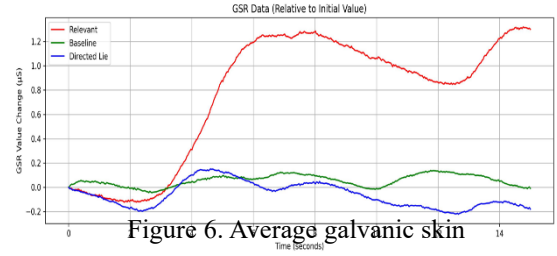


Figure 6. Average galvanic skin

response change curves

Since truth-tellers should react more strongly to comparison questions than to relevant questions, and liars the opposite, a program subtracted the response magnitude of the comparison questions from that of the relevant questions to examine whether the gap between the two question types differs significantly between truth-tellers and liars (truth-tellers tending toward negative values, liars toward positive values). Response magnitude was calculated using the following two methods:

- (1) Maximum-minus-minimum method: for the 0–8 second window, the maximum minus the minimum of pupil diameter change and galvanic skin response was computed for the relevant and comparison questions, representing the response magnitude when the examinee was tested on each question type. With D denoting the response-magnitude gap, R the relevant question, and C the comparison question, the formula is visualized as follows

$$D = R_{Max-Min} - C_{Max-Min}$$

- (2) Area-under-curve subtraction method (AUC method): the areas under the curves of pupil diameter change and galvanic skin response in the 0–8 second window were computed for the relevant and comparison questions and subtracted, representing the examinee's response magnitude for the two question types

$$D_{AUC} = R_{AUC} - C_{AUC}$$

2. Unweighted Accuracy, Sensitivity, Specificity, Positive Predictive Value (PPV), and Negative Predictive Value (NPV)

The experimental results were evaluated both by the MSRD system and by manual interpretation by a polygraph examiner, and the two sets of results were compared. After the determinations of deception indicated and no deception indicated were obtained, they were compared against the ground truth of the experimental design to obtain the counts of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Unweighted accuracy, sensitivity, specificity, PPV, and NPV were computed with the following formulas:

$$\text{Unweighted Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Sensitivity} = \frac{TP}{TP+FN}$$

$$\text{Specificity} = \frac{TN}{TN+FP}$$

$$\text{PPV} = \frac{TP}{TP+FP}$$

$$\text{NPV} = \frac{TN}{TN+FN}$$

(1) MSRD System Determination

The MSRD analysis software designed by the Singapore instrument company AISeer averages the subject's responses to the relevant and comparison questions across the three tests, plots the response curves, and computes the response difference in the 0–8 second window, analyzing it with the software's built-in classification module. Figure 7 below shows an example of the software's determination results:

Alibi Game Results

Statements

Statement 1:

你去派對的事,你有沒有說謊?

Predictions

GSR: Lie , 52.30%

Overall: Lie , 57.00%

Pupil: Lie , 58.00%

Compliance Metrics

Compliance Directed Lie:

6600.00%

Questions

Question 1: 你去派對的事,你有沒有說謊?

Answer 1: NO

Question 2: 你週六去派對的事,你有沒有說謊?

Answer 2: NO

Question 3: 你週六去台北派對的事,你有沒有說謊?

Answer 3: NO

Figure 7. Example of MSRD software

~~As the figure~~ determination results

shows, the software is designed to label a physiological channel “Lie” when its response proportion exceeds 50% and “Truth” at 50% or below; GSR denotes the galvanic skin response, Pupil denotes pupil diameter change, and Overall is the combined result of the two.

(2) Manual Interpretation by a Polygraph Examiner

In this part, an examiner with 19 years of polygraph experience performed chart analysis on the average response curves for the relevant and comparison questions of the three charts per test provided by AISeer, likewise assessing the highest and lowest points of the subject’s galvanic skin response and pupil diameter change within 0–8 seconds, and scoring with a three-position method: if the difference between the highest and lowest points was greater for the comparison question than for the relevant question, +1 was given; otherwise –1; when manual judgment could not discern a difference or the chart was disordered, 0 was given. Examples are shown in Figures 8 to 9

Average Pupil Response Post Question Stimuli

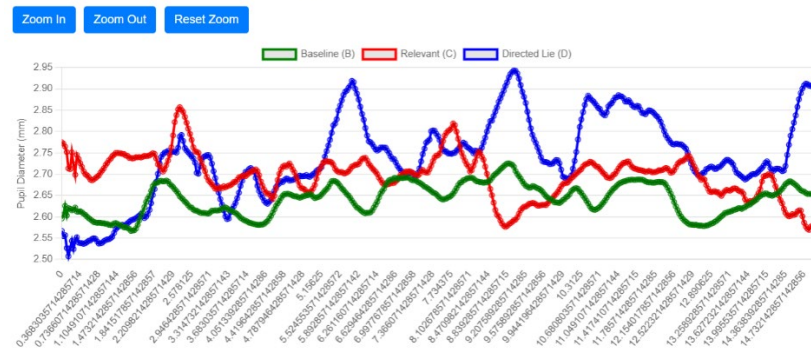


Figure 8. Example chart for manual interpretation (pupil diameter change;

this chart scores +1)

Average GSR Response Post Question Stimuli

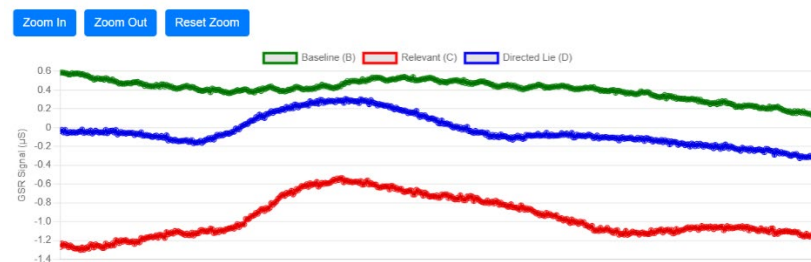


Figure 9. Example chart for manual interpretation (galvanic skin response

change; this chart scores -1)

The galvanic skin response and pupil diameter change scores of the three physiological charts were each summed, and the final total was computed, to analyze the lie-discrimination ability of galvanic skin response and pupil diameter change individually as well as combined. For the decision threshold, a total of -1 or below was classified as deception indicated, $+1$ or above as no deception indicated, and 0 as inconclusive.

4. Results and Discussion

- (1) Results of differences in pupil diameter change and galvanic skin response within the response window and statistical tests

1. Results of the maximum-minus-minimum method

Table 4.1.1.1 Maximum-minus-minimum differences in pupil diameter change for the deceptive and truthful groups

	Mean (SD)	t value	Cohen's d
Deceptive group	-0.036 (.089)	2.432*	0.55
Truthful group	-0.092 (.114)		

N=80, * $p < .05$

Table 4.1.1.2 Maximum-minus-minimum differences in galvanic skin response for the deceptive and truthful groups

	Mean (SD)	t value	Cohen's d
Deceptive group	.187 (.347)	3.347*	0.75
Truthful group	-.168 (.573)		

N=80, * $p < .05$

2. Results of the AUC method

Table 4.1.2.1 AUC differences in pupil diameter change for the deceptive and truthful groups

	Mean (SD)	t value	Cohen's d
Deceptive group	-.282 (.668)	2.610*	0.58
Truthful group	-.748 (.910)		

N=80, * $p < .05$

Table 4.1.2.2 AUC differences in galvanic skin response for the deceptive and truthful groups

	Mean (SD)	t value	Cohen's d
Deceptive group	.574 (1.402)	2.610*	0.83
Truthful group	-1.153 (2.573)		

N=80, * $p < .05$

(2) Confusion matrices and unweighted accuracy

1. MSRD system determination: confusion matrix and accuracy

Ground-truth group	Test result	
	DI	NDI
Deceptive group	32	8
Truthful group	16	24

Note 1: N=80

Note 2: In the test-result columns, DI means Deception Indicated and NDI means No Deception Indicated

$$\text{Unweighted Accuracy} = \frac{32+24}{80} \times 100\% = 70\%$$

$$\text{Sensitivity} = \frac{32}{40} = 0.80$$

$$\text{Specificity} = \frac{24}{40} = 0.60$$

$$\text{Positive predictive value} = \frac{32}{48} = 0.67$$

$$\text{Negative predictive value} = \frac{24}{32} = 0.75$$

2. Manual analysis of MSRD charts: confusion matrix and accuracy

Ground-truth group	Test result	
	DI	NDI
Deceptive group	27	8
Truthful group	6	31

Note 1: N=72, INC=8

Note 2: In the test-result columns, DI means Deception Indicated and NDI means

No Deception Indicated

$$\text{Unweighted Accuracy} = \frac{27+31}{72} \times 100\% = \mathbf{80.56\%}$$

$$\text{Sensitivity} = \frac{27}{35} = \mathbf{0.77}$$

$$\text{Specificity} = \frac{31}{37} = \mathbf{0.84}$$

$$\text{Positive predictive value} = \frac{27}{33} = \mathbf{0.82}$$

$$\text{Negative predictive value} = \frac{31}{39} = \mathbf{0.79}$$

3. MSRD system, pupil diameter change: confusion matrix and accuracy

Ground-truth group	Test result	
	DI	NDI
Deceptive group	17	23
Truthful group	6	34

Note 1: N=80

Note 2: In the test-result columns, DI means Deception Indicated and NDI means No Deception Indicated

$$\text{Unweighted Accuracy} = \frac{17+34}{80} \times 100\% = \mathbf{63.75\%}$$

$$\text{Sensitivity} = \frac{17}{40} = \mathbf{0.43}$$

$$\text{Specificity} = \frac{34}{40} = \mathbf{0.85}$$

$$\text{Positive predictive value} = \frac{17}{23} = \mathbf{0.74}$$

$$\text{Negative predictive value} = \frac{34}{57} = \mathbf{0.60}$$

4. Manual interpretation of pupil diameter change in MSRD charts: confusion matrix and accuracy

Ground-truth group	Test result	
	DI	NDI
Deceptive group	15	20
Truthful group	4	34

Note 1: N=73, INC=7

Note 2: In the test-result columns, DI means Deception Indicated and NDI means No Deception Indicated

$$\text{Unweighted Accuracy} = \frac{15+34}{73} \times 100\% = \mathbf{67.12\%}$$

$$\text{Sensitivity} = \frac{15}{35} = \mathbf{0.43}$$

$$\text{Specificity} = \frac{34}{38} = \mathbf{0.89}$$

$$\text{Positive predictive value} = \frac{15}{19} = \mathbf{0.79}$$

$$\text{Negative predictive value} = \frac{34}{54} = \mathbf{0.63}$$

5. MSRD system, galvanic skin response: confusion matrix and accuracy

Ground-truth group	Test result	
	DI	NDI
Deceptive group	27	13
Truthful group	14	26

Note 1: N=80

Note 2: In the test-result columns, DI means Deception Indicated and NDI means No Deception Indicated

$$\text{Unweighted Accuracy} = \frac{27+26}{80} \times 100\% = \mathbf{66.25\%}$$

$$\text{Sensitivity} = \frac{27}{40} = \mathbf{0.68}$$

$$\text{Specificity} = \frac{26}{40} = \mathbf{0.65}$$

$$\text{Positive predictive value} = \frac{27}{41} = \mathbf{0.66}$$

$$\text{Negative predictive value} = \frac{26}{39} = \mathbf{0.67}$$

6. Manual interpretation of galvanic skin response in MSRD charts: confusion matrix and accuracy

Ground-truth group	Test result	
	DI	NDI
Deceptive group	29	5
Truthful group	10	23

Note 1: N=67, INC=13

Note 2: In the test-result columns, DI means Deception Indicated and NDI means No Deception Indicated

$$\text{Unweighted Accuracy} = \frac{29+23}{67} \times 100\% = \mathbf{77.61\%}$$

$$\text{Sensitivity} = \frac{29}{34} = \mathbf{0.85}$$

$$\text{Specificity} = \frac{23}{33} = \mathbf{0.70}$$

$$\text{Positive predictive value} = \frac{29}{39} = \mathbf{0.74}$$

$$\text{Negative predictive value} = \frac{23}{28} = \mathbf{0.82}$$

7. Lafayette Instrument EDA accuracy

Ground-truth group	Test result	
	DI	NDI
Deceptive group	11	6
Truthful group	0	16

Note 1: N=33, INC=2

Note 2: In the test-result columns, DI means Deception Indicated and NDI means No Deception Indicated

$$\text{Unweighted Accuracy} = \frac{11+16}{33} \times 100\% = \mathbf{81.82\%}$$

$$\text{Sensitivity} = \frac{11}{17} = \mathbf{0.65}$$

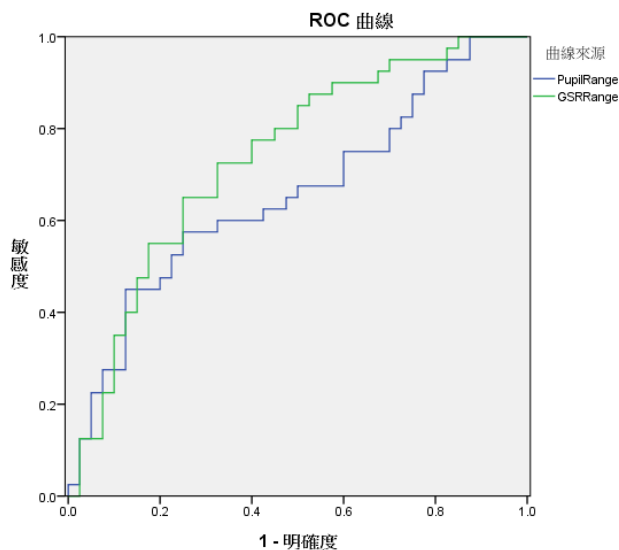
$$\text{Specificity} = \frac{16}{16} = \mathbf{1.00}$$

$$\text{Positive predictive value} = \frac{11}{11} = \mathbf{1.00}$$

$$\text{Negative predictive value} = \frac{16}{22} = \mathbf{0.73}$$

(3) ROC curve comparison of pupil diameter change and galvanic skin response

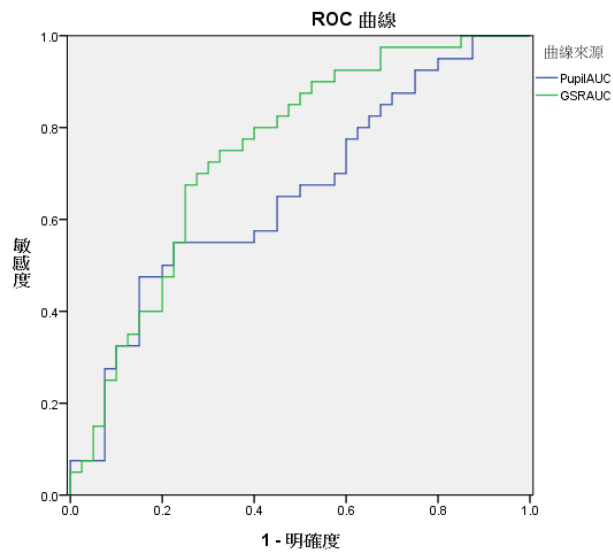
1. Comparison using the maximum-minus-minimum difference



The AUC of the ROC curve was .656 for pupil diameter change and .737

for galvanic skin response, indicating that the galvanic skin response discriminates deception better than pupil diameter change

2. Comparison using the AUC method



The AUC of the ROC curve was .658 for pupil diameter change and .747 for galvanic skin response, indicating that the galvanic skin response discriminates deception better than pupil diameter change

5. Conclusions

- (1) For the statistical tests of the relevant-minus-comparison differences in pupil diameter change and galvanic skin response, whether analyzed by the maximum-minus-minimum method or the AUC method, the deceptive and truthful groups differed significantly in both physiological channels. Observing the means, the deceptive group's relevant-minus-comparison differences were larger than the truthful group's in every case, supporting Hypothesis 1.
- (2) Regarding accuracy, although the MSRD system's accuracy when combining pupil diameter change and galvanic skin response was only 70%, expert manual chart analysis—incorporating inconclusive determinations—achieved an unweighted accuracy of 80.56%, better than pupil diameter change alone (63.75%) or galvanic skin response alone (66.25%), and better than the system's combined determination (70%), supporting Hypothesis 2. In expert manual interpretation, the accuracy of pupil diameter change was 67.12% and that of galvanic skin response 77.61%; for each individual channel, expert manual interpretation outperformed the system's determination.
- (3) Because only the expert manual results incorporated inconclusive determinations, Hypothesis 3 is discussed only with respect to the expert manual results. Using galvanic skin response alone, expert manual interpretation gave an unweighted accuracy of 77.61% and an inconclusive rate of 16.25% (13/80); when combined with pupil diameter change, the unweighted accuracy and inconclusive rate were 80.56% and 10%, respectively. This not only supports Hypothesis 3 but also satisfies APA Standard of Practice 1.1.7.3.3, which requires investigative PDD examinations to average at least 80% accuracy, and item 6 of ASTM E1954-05, which requires a utility value of no less than 0.8.
- (4) In the comparison question technique, pupil diameter change shows low sensitivity and high specificity, echoing the study of the Utah comparison question technique conducted by Webb, Honts, et al. (2009) at the University of Utah. Webb, Honts, et al. (2009) concluded that innocent subjects' pupil dilation was significantly greater for comparison questions than for relevant questions, while guilty subjects showed no significant difference between the two question types. They held that lying requires more cognitive load than answering truthfully, and pupil diameter reflects changes in cognitive load; innocent subjects lie only when answering comparison questions, so pupil dilation on

comparison questions should be more pronounced than on relevant questions. On the other hand, guilty subjects showed no significant difference in pupil dilation between comparison and relevant questions, presumably because they had to respond deceptively to both question types, so their pupil responses may simply reflect this. In the present study, deceptive subjects' pupil diameter changes showed no significant difference between the comparison and control questions, which manifested in the test results as inconclusive determinations or low-sensitivity outcomes. Apart from the cognitive-load differences caused by question type, a closer examination of this study's experimental design suggests that it cannot be ruled out that, in the experimental method, requiring subjects to simulate lying about everyday life events imposed a relatively low cognitive load on them; by contrast, the moral directed lie questions, which required subjects to think and then answer in reverse, caused subjects to expend considerable cognitive load, affecting pupil diameter change and preventing the chart results from showing a clear distinction—hence the high-specificity, low-sensitivity result for this physiological channel.

- (5) Lafayette Instrument EDA accuracy slightly higher than MSRD
This paper made an exploratory comparison of the accuracy of the Lafayette Instrument EDA and the MSRD. Although only 35 samples were selected during the later stage of the study—raising a possible concern of deliberate selection—the Lafayette Instrument EDA was not used to collect subjects' galvanic skin responses before September 24, so the results retain reference value; for a more rigorous comparison, a dedicated comparative experiment is recommended. Preliminarily, the Lafayette Instrument EDA performed slightly better than the MSRD.
- (6) Research limitations and issues with the MSRD system and scenario design:
 1. Inability to observe whether subjects used countermeasures
The activity-monitoring seat pad and respiration sensors typically paired with a standard polygraph were not used, so common countermeasures such as a subject randomly moving hands and feet or breathing deeply cannot be reflected in the MSRD system's charts.
 2. The moral directed lie question showed high specificity but poor sensitivity in pupil diameter change; it cannot be ruled out that the experimental scenario, centered on everyday-life memories, weakened

subjects' responses to the relevant questions.

3. The experimental lying scenario was confined to everyday-life memories and lacked the emotional pressure and anxiety of real criminal cases, so its ecological validity in real cases remains to be clarified.
4. The physiological channels used by the experimental equipment included no channel for monitoring countermeasures, so it cannot be confirmed whether the physiological responses were contaminated by countermeasures.

(7) Suggestions for future research

With advances in physiological monitoring technology, the software and hardware for measuring pupil diameter change have improved and become easier to use, greatly reducing measurement barriers for researchers. This study found that although this channel has low sensitivity, pairing it with a high-sensitivity channel allows the two to complement each other. Future research is advised to make the experimental scenario closer to real-case conditions—for example, by using mock crimes or increasing experimental rewards—to strengthen subjects' motivation to be truthful or to evade detection.

References

- [1] Shurany, T., & Gordon, N. J. (2021). Directed Lie—The Correct or the Easy Way? *European Polygraph*, 15(1 (53), 29-38.
- [2] Honts, C. R. (1994). Psychophysiological detection of deception. *Current Directions in Psychological Science*, 3(3), 77-82.
- [3] Krapohl, D., Handler, M., & Lynch, M. (2023). Terminology Reference for the Science of Psychophysiological Detection of Deception 4th Edition, 2022. *European Polygraph*, 17(2 (58), 11-149.
- [4] Iacono, W. G. (2024). Psychology and the lie detector industry: A fifty-year perspective. *Biological Psychology*, 190, 108808.
- [5] Rashid, A. S., Carmichael, B., Su, C., Shi, K., Lim, K., Poorvika, S. K., Wan, N. J., Govil, E., & Yap, D. (2024). Evaluating Physiological Indicators in Detecting Deception and Truthfulness using the Comparison Question Test. *bioRxiv*, 2024.2009.2001.609623.
- [6] Nelson, R. (2015). Scientific basis for polygraph testing. *Polygraph*, 44(1), 28-61.
- [7] Senter, S., Weatherman, D., Krapohl, D., & Horvath, F. (2010). Psychological set or differential salience: A proposal for reconciling theory and terminology in polygraph testing. *Polygraph*, 39(2), 109-117.
- [8] Kahn, J., Nelson, R., & Handler, M. (2009). An exploration of emotion and cognition during polygraph testing. *Polygraph*, 38(3), 184-197.
- [9] Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, 154(3756), 1583-1585.
- [10] Lubow, R., & 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.
- [11] Webb, A. K., Honts, C. R., Kircher, J. C., Bernhardt, P., & Cook, A. E. (2009). Effectiveness of pupil diameter in a probable-lie comparison question test for deception. *Legal and Criminological Psychology*, 14(2), 279-292.
- [12] Cook, A. E., Potts, A., Kupcova, M., Hacker, D. J., Woltz, D. J., & Kircher, J. C. (2024). Using pupillometry in the detection of deception. In *Modern Pupillometry: Cognition, Neuroscience, and Practical Applications* (pp. 349-372). Springer.
- [13] Johnson Jr, R., Barnhardt, J., & Zhu, J. (2005). Differential effects of practice on the executive processes used for truthful and deceptive responses: An event-related brain potential study. *Cognitive Brain Research*, 24(3), 386-404.
- [14] Kircher, J. (1981). Psychophysiological processes in the detection of deception. *Unpublished doctoral preliminary examination. Department of Psychology, University of Utah.*
- [15] Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities*. John Wiley & Sons.
- [16] Walczyk, J. J., Mahoney, K. T., Doverspike, D., & Griffith-Ross, D. A. (2009). Cognitive lie detection: Response time and consistency of answers as cues to deception. *Journal of Business and Psychology*, 24(1), 33-49.

- [17] Walczyk, J. J., Roper, K. S., Seemann, E., & Humphrey, A. M. (2003). Cognitive mechanisms underlying lying to questions: Response time as a cue to deception. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 17(7), 755-774.
- [18] Ahern, S., & Beatty, J. (1979). Pupillary responses during information processing vary with Scholastic Aptitude Test scores. *Science*, 205(4412), 1289-1292.
- [19] Beatty, J., & Wagoner, B. L. (1978). Pupillometric signs of brain activation vary with level of cognitive processing. *Science*, 199(4334), 1216-1218.
- [20] Colman, F., & Paivio, A. (1970). Pupillary dilation and mediation processes during paired-associate learning. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 24(4), 261.
- [21] Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, 143(3611), 1190-1192.
- [22] Dionisio, D. P., Granholm, E., Hilix, W. A., & Perrine, W. F. (2001). Differentiation of deception using pupillary responses as an index of cognitive processing. *Psychophysiology*, 38(2), 205-211.
- [23] Bradley, M., & Janisse, M. P. (1981). Accuracy demonstrations, threat, and the detection of deception: Cardiovascular, electrodermal, and pupillary measures. *Psychophysiology*, 18(3), 307-315.
- [24] Heilveil, I. (1976). Deception and pupil size. *Journal of Clinical Psychology*, 32(3).
- [25] Webb, A. K., Hacker, D. J., Osher, D., Cook, A. E., Woltz, D. J., Kristjansson, S., & Kircher, J. C. (2009). Eye movements and pupil size reveal deception in computer administered questionnaires. International conference on foundations of augmented cognition,
- [26] Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American educational research journal*, 45(1), 166-183.
- [27] Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. In *Handbook of metacognition in education* (pp. 299-315). Routledge.
- [28] Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Roth, W. T., Dawson, M. E., & Filion, D. L. (2012). Society for psychophysiological research ad hoc committee on electrodermal measures. Publication recommendations for electrodermal measurements. *Psychophysiology*, 49(8), 1017-1034.
- [29] Rahma, O. N., Putra, A. P., Rahmatillah, A., Putri, Y. S. a. K. A., Fajriaty, N. D., Ain, K., & Chai, R. (2022). Electrodermal activity for measuring cognitive and emotional stress level. *Journal of Medical Signals & Sensors*, 12(2), 155-162.
- [30] Meijer, E. H., Selle, N. K., Elber, L., & Ben-Shakhar, G. (2014). Memory detection with the Concealed Information Test: A meta analysis of skin conductance, respiration, heart rate, and P300 data. *Psychophysiology*, 51(9), 879-904.
- [31] Nelson, R. (2017). Multinomial reference distributions for the Empirical Scoring System. *Polygraph & Forensic Credibility Assessment*, 46(2), 81-115.

- [32] Menges, P. M. (2004). Directed lie comparison questions in polygraph examinations: history and methodology. *Polygraph*, 33(3), 131-142.
- [33] Honts, C. R., & Raskin, D. C. (1988). A field study of the validity of the directed lie control question. *Journal of Police Science and Administration*, 16(1), 56-61.
- [34] Horowitz, S. W., Kircher, J. C., Honts, C. R., & Raskin, D. C. (1997). The role of comparison questions in physiological detection of deception. *Psychophysiology*, 34(1), 108-115.
- [35] Kahneman, D. (2013). *A perspective on judgment and choice: Mapping bounded rationality*. Psychology Press.
- [36] Stanovich, K. E. (1999). *Who is rational?: Studies of individual differences in reasoning*. Psychology Press.
- [37] Kvaran, T., Nichols, S., & Sanfey, A. (2013). The effect of analytic and experiential modes of thought on moral judgment. *Progress in brain research*, 202, 187-196.