Effects of the combination of P3-based GKT and reality monitoring on deceptive classification

Ki-Won Jang[†], Deok-Yong Kim[†], Sungkun Cho and Jang-Han Lee*

Clinical Neuro-psychology Laboratory, Department of Psychology, Chung-Ang University, Seoul, South Korea

The study aimed to investigate whether a combination of the P3-based Guilty Knowledge Test (GKT) and reality monitoring (RM) distinguished between individuals who are guilty, witnesses, or informed, and using both tests provided more accurate information than did the use of either measure alone. Participants consisted of 45 males that were randomly and evenly assigned to three groups (i.e., guilty, witness, and informed). The guilty group conducted a mock crime where they intentionally crashed their vehicle into another vehicle in a virtual environment (VE). As those in the witness group drove their own vehicles, they observed the guilty groups' vehicle crash into another vehicle. The informed group read an account and saw screenshots of the accident. All participants were instructed to insist that they were innocent. Subsequently, they performed the P3-based GKT and wrote an account of the accident for the RM analysis. A higher P3 amplitude corresponded to how well the participants recognized the presented stimulus, and a higher RM score corresponded to how well the participants reported vivid sensory information and how much less they reported uncertain information. Findings for the P3-based GKT indicated that the informed group showed lower P3 amplitude when presented with the probe stimulus than did the guilty and witness groups. Regarding the RM analysis, the informed group obtained higher RM scores on visual, temporal, and spatial details and lower scores on cognitive operations than the guilty and witness groups. Finally, discriminant analysis revealed that the combination of the P3-based GKT and RM more accurately distinguished between the three groups than the use of either measure alone. The findings suggest that RM may build upon a weakness of the P3-based GKT's. More specifically, it may build upon its susceptibility to the leakage of information about the crime, therefore helping protect innocent individuals who have information about a crime from being perceived as guilty.

Keywords: lie detection, Guilty Knowledge Test, reality monitoring, P3, leakage of knowledge

INTRODUCTION

Deception occurs in a variety of interpersonal situations. Individuals are able to detect deception using several methods. When an individual tells a lie, that person unconsciously displays potential cues that may be behavioral, verbal, and psychophysiological (Vrij, 2000; Gamer et al., 2006). Typically, lie detection tools are designed to detect a lie using these cues. One commonly used lie detection tool is the arousal-based polygraph (Vrij, 2000). The arousal-based polygraph detects a lie based on differences in psychophysiological responses (e.g., electrodermal response, cardiovascular, and respiratory) to crime-relevant and crime-irrelevant questions (Kircher and Raskin, 1988; Richardson et al., 1990; Ben-Shakhar et al., 1999). Despite its usefulness in the field, the arousal-based polygraph has some limitations. For example, the arousal-based polygraph indirectly detects a lie by measuring variables related to the lie (e.g., guilty, anxiety). However, these emotions can appear not only in a situation where individuals tell a lie but also in an uncertain situation where innocent individuals are false accused (Allen and Mertens, 2009).

Another method of lie detection is based on recognition of crime-related information that is stored in memory. This method

is called the Guilty Knowledge Test (GKT). Individuals who commit a crime have specific knowledge or memories about the crime, whereas innocent individuals do not. The GKT examines whether suspects possess this specific knowledge. If a guilty suspect recognizes the crime-related evidence presented, he or she is more likely to produce higher physiological responses than will non-guilty suspects (Vrij, 2000). One of the methods used to detect such recognition is the event-related potentials (ERPs). The ERPs provides considerably accurate information regarding temporal changes in brain activity in response to the processing of a particular stimulus. Of the ERPs components, the P300 (P3) component is evoked in response to attentive, recognized, and meaningful stimuli (Polich and Kok, 1995; Polich, 2000).

The P3-based GKT is a tool that uses temporal changes in brain activity to detect deception (Farwell and Donchin, 1991; Rosenfeld et al., 1991; Allen et al., 1992; Abootalebi et al., 2006). In the P3-based GKT, three types of stimuli, target, probe, and irrelevant, are presented to participants. Probe stimuli contain concealed crime-related information that only the guilty individuals possess (Ben-Shakhar and Elaad, 2002). Irrelevant stimuli are

stimuli with information that is unrelated to the crime. Target stimuli are also stimuli that contain information unrelated to the crime. Target stimuli are essentially a list of stimuli the participants are instructed to respond to by pressing a button when the stimuli are presented (Farwell and Donchin, 1991). The basic assumption of the P3-based GKT is that guilty individuals will recognize the probe stimuli, thus evoking higher P3 amplitude in their brain potentials compared to that evoked by the irrelevant stimuli. Conversely, innocent individuals will show no differences in P3 amplitude in brain potentials in response to the probe and irrelevant stimuli. Prior studies of ERPs-based GKT demonstrated that accurate rates of deception detection are relatively high, ranging from 89 to 90% (Farwell and Donchin, 1991; Rosenfeld et al., 1991; Allen et al., 1992). Despite its usefulness, the P3-based GKT is weak in natural environments. For example, probe stimuli may be disclosed to the public through mass media or disseminated verbally by individuals participating in the investigation (e.g., a criminal investigator). The witnesses may also have information about the crime even though they did not conduct a crime. Thus, they may not be classified into the innocent groups because of their knowledge or memories of the crime, which discourages the use of the P3-based GKT (Ben-Shakhar et al., 1999).

The quality of information about the crime may be different for guilty vs. innocent individuals even though they both have information about a particular crime. Given that they conducted the crime, guilty individuals may have more vivid sensory information about the crime as compared to innocent individuals. Such differences can be revealed by reality monitoring (RM), one type of statement analyses conducted in a crime setting. RM is based on differences in the quality of information contained in an individual's memory for experienced vs. imagined events (Johnson and Raye, 1981; Memon et al., 2010). The assumption of RM is that experienced memories differ from fictional memories (e.g., Vrij et al., 2001). The RM evaluates the quality of information contained in memories using several criteria, including visual, auditory, temporal, and spatial details as well as cognitive operations. The visual and auditory details involve perceptual and sensory information. The temporal details provide information about timing or duration of events. The spatial details include information regarding where the event took place and how objects and people were situated in relation to each other. Memories of experienced events may involve more perceptual, sensory, spatial, and temporal details than the memories of imagined events. However, imagined memories that are obtained through an internal source are likely to contain thoughts and reasoning in one's testimony, called cognitive operations. Cognitive operations are usually vague and not concrete (Vrij, 2008) and therefore are less frequent in memories of the experienced events as compared to memories of imagined events (Vrij et al., 2004). In criminal situations, an individual's statement about an experienced event is likely to contain the truth, whereas a statement pertaining to an imagined event is likely to contain false information (Vrij, 2008). Thus, the RM analysis would differentiate statements between those who actually experience a crime (guilty individuals) and those who receive information about the crime only (witness

or informed individuals). By distinguishing between those who have experienced a crime and those who only have information about a crime, RM can determine whether individuals are innocent. The individuals who witnessed the crime are more likely to describe perceptual and sensory details of the experienced event in their statements, whereas the guilty and informed individuals may describe the imagined event without actually experiencing it. For example, the guilty and informed individuals are more likely to make up a story either because they did not actually experience the crime or because had to pretend being innocent. Given these, RM may be used to complement and overcome a possible weakness of the P3-based GKT.

The aim of this study was to investigate whether a combination of the P3-based GKT and RM would more effectively differentiate among guilty individuals who conduct a crime, witnesses who experience the crime, and informed individuals who did not experience the crime but have information about the crime. Several hypotheses guide the current study. First, we predicted that in the P3-based GKT, the guilty group would show higher P3 amplitude in response to the probe stimuli than the other two groups. Second, we predicted that the witness group would meet the RM criteria more frequently than the other two groups. Third, we predicted that the combination of the P3-based GKT and RM would discriminate the groups more accurately than the P3-based GKT or RM alone.

MATERIALS AND METHODS

The participants consisted of 45 male undergraduates (15 per group), and the mean age of the sample was 23.07 years (SD = 2.41). They were randomly and evenly assigned to three groups: guilty, witness, and informed. This study used male participants because gender may affect ERPs amplitude (e.g., Cahill and Polich, 1992; Polich and Martin, 1992; Reinvang, 1999). We used the Machiavellianism Scale (Christie and Geis, 1970), the Social Adroitness Scale (Jackson, 1994), and Levenson's Self-Report Psychopathy Scale (LSRP) (Levenson et al., 1995) to control for the manipulativeness of the participants. No significant differences were found among groups; the Machiavellianism Scale $F_{(2, 42)} = 0.68$, n.s.; the Social Adroitness Scale $F_{(2, 42)} = 0.43$, n.s.; the LSRP $F_{(2, 42)} = 0.44$, n.s. (**Table 1**).

A 3D visual display was presented on dual monitors through an Olympus FMD-250W head-mounted display with a resolution of 800 × 600 pixels. A computer game (Grand Theft Auto

Table 1 | Demographic data and the questionnaire scores.

	Mean (SD)		
	Guilty group (*n* = 15)	Witness group (*n* = 15)	Informed group (*n* = 15)
Age	23.27 (1.94)	22.93 (2.84)	23.00 (2.51)
Machiavellianism	50.87 (5.18)	51.13 (6.10)	48.80 (6.68)
Social Adroitness	12.60 (2.32)	11.67 (3.39)	11.87 (2.85)
LSRP	36.47 (6.97)	37.60 (6.58)	35.40 (5.65)

LSRP, Levenson's self-report psychopathy scale.

San Andreas: GTASA) that involved a third-person shooter and a driving simulator in a virtual environment (VE) was used.

When the participants arrived at the laboratory, they were asked to sign a consent form and then the experimenter explained the objectives of the study. Each participant was randomly assigned to one of the groups (i.e., guilty, witness, or informed). The guilty or witness group wore a head-mounted device while experiencing the VE. Two vehicles were driven in the VE. One vehicle was driven by the guilty or witness group, and the other one was driven by an experimenter. The guilty group intentionally crashed their vehicle into the experimenter's vehicle (i.e., head-on collision). The crash was severe enough to blow the hood off the vehicles. The witness group was instructed to drive safely and watch the crash that was caused by the experimenter. The informed group was instructed to watch screenshots of the car accident caused by the guilty group and read the description of the accident. The duration of the VE was approximately 15 min. All of the groups were instructed to write statements about the car accident and to insist that they were innocent.

Subsequently, all participants were given the GKT to evaluate whether they recognized the vehicle of assailant. Three types of stimuli (irrelevant, probe, and target stimuli) were used in the GKT. The irrelevant stimuli consisted of four screenshots of a vehicle not related to the accident, and the probe stimulus was a screenshot of the assailant's vehicle. All participants were instructed to respond by pressing a spacebar when the target stimulus (i.e., a vehicle unrelated to the car accident) was presented. The vehicles used in this experiment were similar in size and color but different in shape. Each trial consisted of one probe, four irrelevant, and one target stimuli (total 40 trials). After presenting a fixation for 500 ms, the probe, irrelevant, and target stimuli were randomly presented for 1000 ms, and then a blank screen followed for 1500 ms with an inter-stimulus-interval (ISI) of 3000 ms (visual angle being 16° in width and 13° in height; 518 × 370 pixels). After attaching the electrodes to the head of each participant, the participants were instructed to press a "yes" button when the target stimuli were presented and a "no" button when the others were presented. The total duration of the P3-based GKT was approximately 12 min.

EEG data were recorded from 28 sites (Fpz, Fz, FCz, Cz, CPz, Pz, Oz, Fp2, F3/4, F7/8, FC3/4, C3/4, CP3/4, P3/4, P7/8, O1/2, T7/8, and FT7/8) with reference electrodes on the earlobes using the Neuroscan System (Neuroscan Labs, Sterling, VA, USA). EOGs were recorded from the outer canthus of each eye to measure horizontal electro-oculograms (HEOG) activity and from the left eye to measure vertical electro-oculograms (VEOG) activity. All impedances were maintained at 5 kΩ or less. The data were digitized at a rate of 512 Hz for 800 ms and recorded with a band-pass of 0.01–100 Hz. Epochs were created from -100 to 898 ms around the stimulus and baseline corrected using the 100 ms pre-stimulus period. Artifacts in which the EEG or EOG exceeded ±100 μV were rejected (Semlitsch et al., 1986). The bandpass filter was applied 0.05–10 Hz (24 dB octave/slope). The P3 component was typically defined as the largest positive peak occurring between 300 and 1000 ms at each electrode (Abootalebi et al., 2006). Amplitude was measured as the difference between the mean pre-stimulus baseline and maximum peak amplitude at Pz,

because the Pz site is usually reported to be maximal among the other sites. Peak detection was done automatically but verified manually.

The RM analysis consisted of five domains: visual, auditory, temporal, and spatial details as well as cognitive operations. First, a statement met the criteria of having visual details if it contained a vivid or clear description. Second, a statement met the criteria for having auditory details if it encompassed auditory information. Third, a statement met the criteria for temporal details if it included the order in which the accident occurred. Fourth, a statement met the criteria for spatial details if the statement encompassed locational information on humans or objects. Fifth, a statement met the criteria for cognitive operations if it contained descriptions of imagined events from internal source, such as thoughts and reasoning. Cognitive operations were scored dichotomously: a score of 0 when cognitive operations were not present and a score of 1 when cognitive operations were presented. For the RM, a score for each domain was calculated by summing of frequency of meeting the criteria in their statements by two independent raters (Vrij et al., 2004).

RESULTS

P3-BASED GKT

A 3 (group: guilty, witness, and informed) × 3 (stimulus: probe, irrelevant, and target) repeated-measures ANOVA was performed to analyze the P3 data. The results indicated that there was a significant interaction effect between group and stimulus, $F_{(4, 84)} = 3.27$, $p < 0.05$, $\eta^2 = 0.14$. Subsequently, we performed One-Way ANOVA for each stimulus to investigate differences among the groups. There was a significant main effect of group membership on the P3 amplitudes for the probe stimuli, $F_{(2, 42)} = 8.42$, $p = 0.001$, $\eta^2 = 0.29$, but there were no significant group differences for the target and irrelevant stimuli, $F_{(2, 42)} = 0.31$; $F_{(2, 42)} = 0.37$, n.s. In the pairwise comparison test, the informed group showed lower P3 amplitudes in response to the probe stimuli than the guilty, $t_{(28)} = 4.13$, $p < 0.001$, and witness groups, $t_{(28)} = 2.68$, $p = 0.01$. The difference between the guilty and the witness groups, however, did not reach statistical significance, $t_{(28)} = -1.33$, n.s. (**Figure 1**). Additionally, there was a significant main effect of stimulus, $F_{(2, 84)} = 49.48$, $p < 0.001$, $\eta^2 = 0.54$. A pairwise comparison test indicated that P3 amplitudes in response to the target stimuli were higher than those in response to the probe, $t_{(44)} = 4.43$, $p < 0.001$, and irrelevant stimuli, $t_{(44)} = 8.99$, $p < 0.001$, and that the P3 amplitudes in response to the probe stimuli were significantly higher than the irrelevant stimuli, $t_{(44)} = 5.89$, $p < 0.001$. However, there was no main effect of group (**Figure 2**).
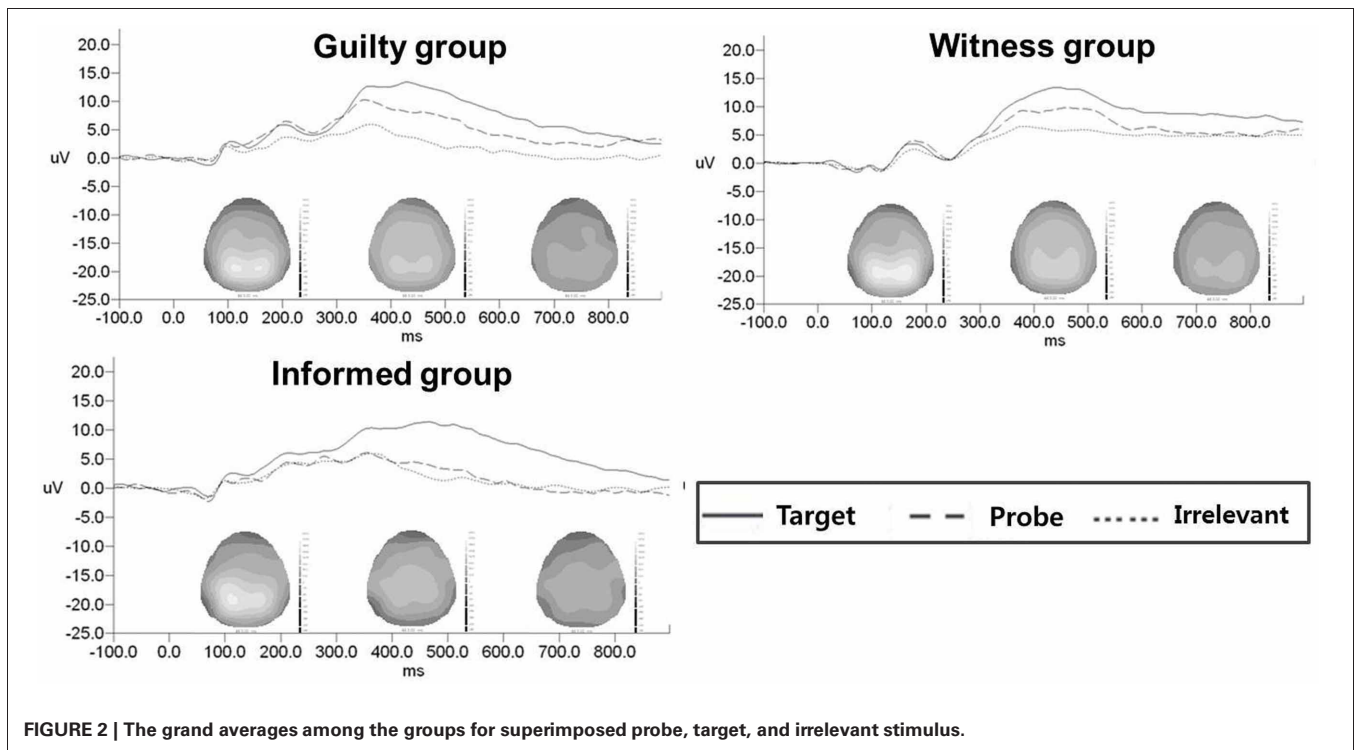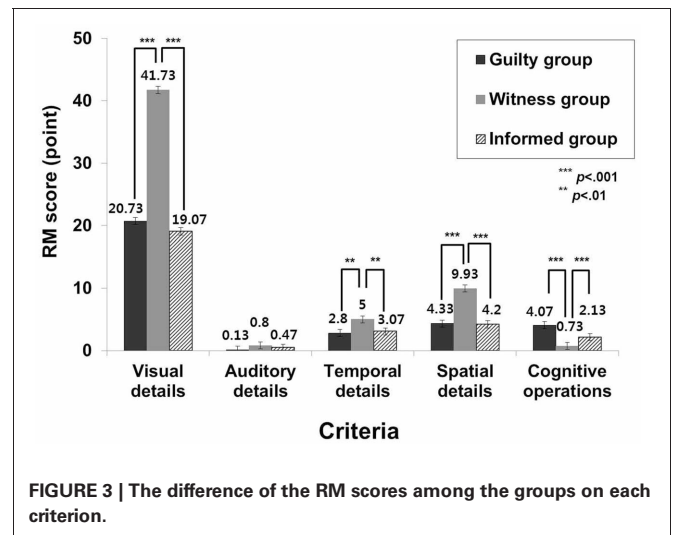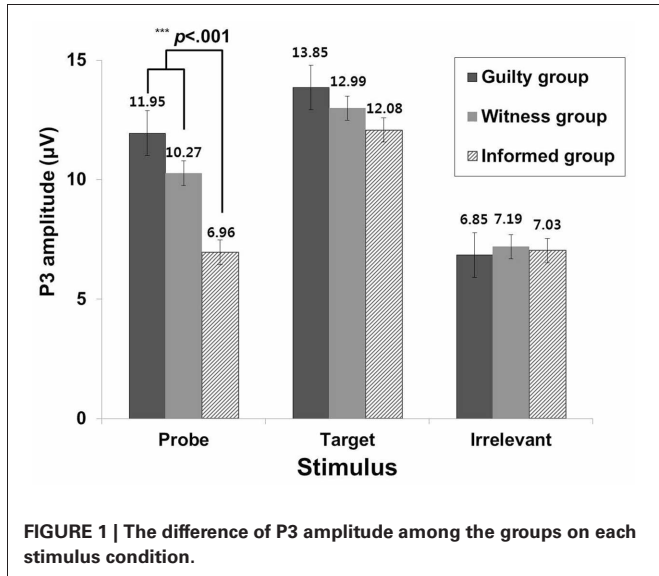
REALITY MONITORING

For the RM analysis, a 2 (group: guilty, witness, and informed) × 5 (criteria: visual, auditory, temporal, spatial, and cognitive operations) MANOVA was used. A MANOVA revealed a significant multivariate main effect for group, Wilks' = 0.13, $F_{(10, 76)} = 13.61$, $\eta^2 = 0.64$. At a univariate level, there were significant main effects for visual, temporal, and spatial details as well as cognitive operations: visual details, $F_{(2, 42)} = 57.10$, $p < 0.001$, $\eta^2 = 0.73$; temporal details, $F_{(2, 42)} = 6.43$, $p < 0.01$,

$\eta^2 = 0.23$; spatial details, $F_{(2, 42)} = 12.57$, $p < 0.001$, $\eta^2 = 0.37$; and cognitive operations, $F_{(2, 42)} = 34.20$, $p < 0.001$, $\eta^2 = 0.62$. A pairwise comparison tests indicated that the witness group reported significantly more visual, temporal, and spatial details than did the guilty [visual details: $t_{(28)} = 8.70$, $p < 0.001$; temporal details: $t_{(28)} = 2.84$, $p < 0.01$; spatial details: $t_{(28)} = 3.72$, $p = 0.001$] and informed groups [visual details: $t_{(28)} = 8.87$, $p < 0.001$; temporal details: $t_{(28)} = 3.04$, $p < 0.01$; spatial details: $t_{(28)} = 3.80$, $p = 0.001$]. In terms of the auditory details, however, there were no significant differences among the groups.

For cognitive operations, the witness group had significantly less than the guilty, $t_{(28)} = -8.30$, $p < 0.001$, and informed group, $t_{(28)} = -3.55$, $p = 0.001$. Furthermore, the informed group also reported cognitive operations significantly less than the guilty group, $t_{(28)} = -4.63$, $p < 0.001$ (**Figure 3**).

DISCRIMINANT ANALYSES FOR P3-BASED GKT AND REALITY MONITORING

A discriminant analysis was conducted to investigate whether a combination of the P3-based GKT and RM was better at distinguishing between the guilty, witness, and informed groups than

FIGURE 1 | The difference of P3 amplitude among the groups on each stimulus condition.

FIGURE 3 | The difference of the RM scores among the groups on each criterion.

FIGURE 2 | The grand averages among the groups for superimposed probe, target, and irrelevant stimulus.

the P3-based GKT (i.e., probe, irrelevant, and target stimuli) or RM (i.e., visual, auditory, temporal, and spatial details as well as cognitive operations). To develop an optimum classifier to discriminate among the groups, two discriminant analyses were conducted. One analysis was performed to find a discriminant function to maximize the separation between the guilty and witness groups, and the other analysis was performed to distinguish between the guilty and informed groups.

First, we conducted a discriminant analysis to determine whether the P3-based GKT, RM, and a combination of the P3-based GKT and RM distinguished between the guilty and witness groups. For the results of the P3-based GKT, univariate F tests showed that the discriminant function was not significant, $\chi^2_{(3, N=30)} = 5.20$, $p = 0.16$, and indicated that 82.2% of total variance was not explained, $\lambda = 0.82$. In the RM, univariate F tests indicated significant differences for the visual, auditory, temporal, and spatial details and cognitive operations, $\chi^2_{(5, N=30)} = 49.93$, $p < 0.001$. The mean classification accuracy was 100.0%. **Table 2** shows that 100.0% of the guilty group and 100.0% of witness group were classified correctly in the present study. An internal validation of the discriminant analysis (jack-knife) also indicated 100.0% correct classifications. The result for the combination of P3-based GKT and RM showed significant differences for the predicted variables, $\chi^2_{(8, N=30)} = 48.60$, $p < 0.001$. The mean classification accuracy was 100.0%. In these analyses, 100.0% of the guilty group and 100.0% of the witness group were classified correctly both in the present study and in the internal validation of the discriminant analysis (**Table 2**).

A discriminant analysis was performed to distinguish between the guilty and informed groups in the P3-based GKT, RM, and combination of the two. The discriminant function analysis results for the P3-based GKT were significant, $\chi^2_{(3, N=30)} = 25.53$, $p < 0.001$, and showed an 86.7% overall correct classification accuracy. A total of 86.7% or the guilty group and 86.7% of the informed group were classified correctly. An internal validation of the discriminant analysis also indicated 86.7% correct classifications. For the results of the RM, the discriminant function indicated significant differences, $\chi^2_{(5, N=30)} = 23.57$, $p < 0.001$. The mean classification accuracy was 86.7%. For these analyses, 86.7% of the guilty group and 86.7% of the informed group were classified correctly. An internal validation of the discriminant analysis showed 76.7% correct classifications, and 66.7% of the guilty group and 86.7% of the informed group were correctly classified. The result of a combination between the P3-based GKT and RM showed significant differences in the predicted variables, $\chi^2_{(8, N=30)} = 34.22$, $p < 0.001$. The overall classification accuracy was 93.3%. In this result, 86.7% of the guilty group and 100.0% of the witness group were classified correctly both in the present study. An internal validation of the discriminant analysis was 90.0% (**Table 3**).

DISCUSSION

The purpose of this study was to investigate differences in the P3-based GKT or RM among individuals in the guilty, witness, and informed groups. Additionally, we investigated whether the combination of the P3-based GKT and RM would more accurately discriminate among the groups than either test alone.

The results indicated that the informed group showed lower P3 amplitude in response to the probe stimulus than did the guilty and witness groups. These results partly support the first hypothesis. Indeed, the results suggest that the P3-based GKT may differentiate individuals who do not experience the crime but who have information about the crime (i.e., informed individuals), from those who do experience it (i.e., witnesses, and guilty individuals). The informed individuals may have less specific memories surrounding they crime compared to individuals

Table 2 | Discriminant analyses with P3-based GKT and reality monitoring between the guilty and witness group.

	Cross validated classification rates (original classification)				
	Hit rates			**Eigenvalue**	**λ**
	Guilty group	**Witness group**	**Total**		
P3-GKT	60.0 (73.3)	40.0 (26.7)	73.3 (80.0)	0.22	0.82
RM	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)	6.09	0.14
P3-GKT + RM	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)	6.58	0.13

Table 3 | Discriminant analyses with P3-based GKT and reality monitoring between the guilty and informed group.

	Cross validated classification rates (original classification)				
	Hit rates			**Eigenvalue**	**λ**
	Guilty group	**Informed group**	**Total**		
P3-GKT	86.7 (86.7)	86.7 (86.7)	86.7 (86.7)	1.62	0.38
RM	66.7 (86.7)	86.7 (86.7)	76.7 (86.7)	1.52	0.40
P3-GKT + RM	86.7 (86.7)	93.3 (100.0)	90.0 (93.3)	3.16	0.24

who experience the crime because they did not directly experience the crime. It may be difficult to identify the source of their crime-related knowledge (Zvi et al., 2012). More specifically, the guilty and witness groups recognized the assailants' better than the informed group, presumably because these two groups had direct experiences with the accident. The P3-based GKT, however, did not reveal significant differences between the guilty and witness groups regarding responses to the probe stimuli. Given these findings, the P3-based GKT appears to be weak in terms of its ability to discriminate between groups when knowledge of a crime is leaked (Ben-Shakhar and Elaad, 2003). The witness and guilty groups may have had similar amounts of vivid information about the crime stored in memory. If this is the case, the P3-based GKT may not have be able to differentiate between the guilty and witness groups. Thus, the P3-based GKT may need another tool to overcome this weakness. More specifically, another tool is needed to differentiate between guilty individuals and witnesses.

Regarding the RM, there were significant differences among the groups, in the visual, temporal, and spatial details as well as cognitive operations. The statements from the witness group included more visual details, temporal information, spatial information, and less cognitive operations than those from the guilty and informed groups. These results may be due to differences in the quality of the crime-related information stored in memory (Vrij et al., 2004). Most likely, the witness group recalled the sensory, perceptual, and contextual memories that they experienced in the VE. The guilty and informed groups, however, described a story that they did not experience. More specifically, the guilty group had to provide a false statement. The informed group had to provide factual statements about the accident. Thus, they were less likely to have limited sensory, perceptual, and contextual memories of the accident. These results suggest that RM may differentiate the experienced-driven true memories and false or imagined memories. Given these findings, the RM seems to be weak in terms of its ability to accurately distinguish among the three groups.

The discriminant analyses revealed that the combination of the P3-based GKT and RM showed higher accuracy rates compared to two methods independently. In terms of the results of the P3-based GKT, the discriminate function was unable to distinguish between the guilty individuals and witnesses, whereas it was able to distinguish between the guilty and informed individuals. These results indicate that the P3-based GKT may not differentiate between witnesses and guilty individuals. Thus, the results suggest the P3-based GKT is weak when knowledge of a crime is leaked. In other words, the witnesses may be falsely accused of committing a crime when the P3-based GKT is employed for the purpose of lie detecting. The results pertaining to RM analysis revealed a high discrimination rate for the witness group (100.0% of the witness group). The combination of the P3-based GKT and RM correctly classified 100.0% of the witness group. This result implies that RM may overcome a weakness of the P3-based GKT. Additionally, the discriminant analysis of the P3-based GKT revealed that it was highly able to discriminate between the guilty and informed individuals, whereas the discriminant analysis of the RM showed moderate discrimination (an internal validation of discrimination analysis: 76.7%). These results

highlight a limitation of RM because both the guilty and informed individuals should possess an imagined memory of the crime that is reflected in their statement. Regarding the overall classification rate, the combination of the P3-based GKT and RM also showed a higher rate of classifications than either the P3-based GKT or RM alone, although differences in correct-classification rates among the techniques were not examined. The results of the present study are comparable to a previous study in which the combination of the Criteria-Based Content Analysis and RM correctly classified 80.8% of the participants (Vrij et al., 2000). However, the aim of the previous study was to discriminate between the guilty group and innocent group that had no information about the crime. In the present study, the combined method showed a higher classification rate (an internal validation between the guilty and witness group of 100.0% and an internal validation between the guilty and informed group of 90.0%). By combing the P3-based GKT and RM, each method builds upon the weaknesses of the other method. In conclusion, the present study suggests that the combination of the P3-based GKT and RM may differentiate among individuals who are guilty, witnesses, and informed.

In the present study, the RM was used to build upon a weakness of the P3-based GKT. Although the combination of the P3-based GKT and RM was adequately able to differentiate between the guilty, witness, and informed groups, this combination also has some possible weakness. If the guilty individuals know the RM criteria, they may be able to manipulate the quality of their report by intentionally changing the balance among cognitive operations, visual, auditory, temporal, and spatial details. Therefore, future studies need to identify the optimal combination of the P3-based GKT and other various methods to differentiate between the guilty individuals and the witnesses or informed individuals.

The present study has several implications. First, the present study suggests that a combination of the P3-based GKT and RM may build upon the weakness of the P3-based GKT because the test is susceptible to the leakage of information about the crime. Therefore, the method may help protect innocent individuals from perceived as guilty when they have information about the crime that was disclosed to the public through mass media or by participating in the investigation.

Second, the present study showed that the GKT using the image stimulus that participants experienced in a VE can discriminate between the groups. Previous study of ERP-based deception detection using a mock crime in a VE indicated that the hit rates were quite low (Mertens and Allen, 2008). Possible reason for such low hit rates would be due to the feature of stimulus. For example, they used stimulus consisting of words, but not images. Many studies have suggested the picture superiority effect (Buckner et al., 2000) that pictures are better recalled than words and better recollected when cued with a fragment only (McBride and Dosher, 2002; Cutmore et al., 2009). Given these, it appears to be reasonable to use image stimuli for a GKT in a VE than word stimuli. Although the findings on the detection of deception using the ERPs-based GKT in a VE have been acceptable in a laboratory experiment, we suggest that future research apply the ERPs-based GKT in a real forensic situation. For example, in a

real forensic situation, there could be delay between conducting a crime and assessing deception. However, in the laboratory setting, the deception is assessed right after conducting a mock crime. Therefore, we suggest that future studies compare memories about a crime both immediately after the crime and after a delayed period.

REFERENCES

Abootalebi, V., Moradi, M. H., and Khalilzadeh, M. A. (2006). A comparison of methods for ERP assessment in a P300-based GKT. *Int. J. Psychophysiol.* 62, 309–320.

Allen, J., Iacono, W. G., and Danielson, K. D. (1992). The identification of concealed memories using the event-related potential and implicit behavioral measures: a methodology for prediction in the face of individual differences. *Psychophysiology* 29, 504–522.

Allen, J., and Mertens, R. (2009). Limitations to the detection of deception: true and false recollections are poorly distinguished using an event-related potential procedure. *Soc. Neurosci.* 4, 473–490.

Ben-Shakhar, G., and Elaad, E. (2002). "The Guilty Knowledge Test (GKT) as an application of psychophysiology: future prospects and obstacles," in *Handbook of Polygraph Testing*, ed M. Kleiner (New York, NY: Academic Press), 87–102.

Ben-Shakhar, G., and Elaad, E. (2003). The validity of psychophysiological detection of information with the Guilty Knowledge Test: a meta-analytic review. *J. Appl. Psychol.* 88, 131–151.

Ben-Shakhar, G., Gronau, N., and Elaad, E. (1999). Leakage of relevant information to innocent examinees in the GKT: an attempt to reduce false-positive outcomes by introducing target stimuli. *J. Appl. Psychol.* 84, 651–660.

Buckner, R. L., Logan, J., Donaldson, D. I., and Wheeler, M. E. (2000). Cognitive neuroscience of episodic memory encoding. *Acta Psychol.* 105, 127–139.

Cahill, J. M., and Polich, J. (1992). P300, probability, and introverted/extroverted personality types. *Biol. Psychol.* 33, 23–35.

Christie, R., and Geis, F. (1970). *Studies in Machiavellianism*. New York, NY: Academic Press.

Cutmore, T. R. H., Djakovic, T., Kebbell, M. R., and Shum, D. H. K. (2009). An object cue is more effective than a word in ERP-based detection of deception. *Int. J. Psychophysiol.* 71, 185–192.

Farwell, L. A., and Donchin, E. (1991). The truth will out: interrogative polygraphy (lie detection) with event-related potentials. *Psychophysiology* 28, 531–547.

Gamer, M., Rill, H., Vossel, G., and Gödert, H. W. (2006). Psychophysiological and vocal measures in the detection of guilty knowledge. *Int. J. Psychophysiol.* 60, 76–87.

Jackson, D. N. (1994). *Jackson Personality Inventory-revised Manual*. Port Huron, MI: Sigma Assessment Systems.

Johnson, M. K., and Raye, C. L. (1981). Reality monitoring. *Psychol. Rev.* 88, 67–85.

Kircher, J. C., and Raskin, D. C. (1988). Human versus computerized evaluations of polygraph data in a laboratory setting. *J. Appl. Psychol.* 73, 291–302.

Levenson, M. R., Kiehl, K. A., and Fitzpatrick, C. M. (1995). Assessing psychopathic attributes in a noninstitutional population. *J. Pers. Soc. Psychol.* 68, 151–158.

McBride, D. M., and Dosher, B. A. (2002). A comparison of conscious and automatic memory processes for picture and word stimuli: a process dissociation analysis. *Conscious. Cogn.* 11, 423–460.

Memon, A., Fraser, J., Colwell, K., Odinot, G., and Mastroberardino, S. (2010). Distinguishing truthful from invented accounts using reality monitoring criteria. *Leg. Crim. Psychol.* 15, 177–194.

Mertens, R., and Allen, J. J. B. (2008). The role of psychophysiology in forensic assessments: deception detection, ERPs, and virtual reality mock crime scenarios. *Psychophysiology* 45, 286–298.

Polich, J. (2000). "P300 in clinical applications," in *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields, 4th Edn.* eds E. Niedermeyer, and F. Lopes Da Silva (Baltimore, Maryland: Lippincott Williams and Wilkins), 1073–1091.

Polich, J., and Kok, A. (1995). Cognitive and biological determinants of P300: an integrative review. *Biol. Psychol.* 41, 103–146.

Polich, J., and Martin, S. (1992). P300, cognitive capability and personality: a correlational study of university undergraduates. *Pers. Individ. Dif.* 13, 533–543.

Reinvang, I. (1999). Cognitive event-related potentials in neuropsychological assessment. *Neuropsychol. Rev.* 9, 231–248.

Richardson, D. C., Carlton, B. L., and Dutton, D. W. (1990). Blind analysis of skin conductance response (SCR) recordings from a number test. *Polygraph* 19, 9–20.

Rosenfeld, J. P., Angell, A., Johnson, M., and Qian, J. H. (1991). An ERP-based, control-question lie detector analog: algorithms for discriminating effects within individuals' average waveforms. *Psychophysiology* 28, 319–335.

Semlitsch, H. V., Anderer, P., Schuster, P., and Presslich, O. (1986). A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP. *Psychophysiology* 23, 695–703.

Vrij, A. (2000). *Detecting Lies and Deceit: The Psychology of Lying and the Implications for Professional Practice*. Chichester: John Wiley and Sons.

Vrij, A. (2008). Nonverbal dominance versus verbal accuracy in lie detection: a plea to change police practice. *Crim. Justice Behav.* 35, 1323–1336.

Vrij, A., Akehurst, L., Soukara, S., and Bull, R. (2004). Let me inform you how to tell a convincing story: CBCA and reality monitoring scores as a function of age, coaching and deception. *Can. J. Behav. Sci.* 36, 113–126.

Vrij, A., Edward, K., and Bull, R. (2001). Stereotypical verbal and nonverbal responses while deceiving others. *Pers. Soc. Psychol. Bull.* 27, 899–909.

Vrij, A., Edward, K., Roberts, K. P., and Bull, R. (2000). Detection deceit via analysis of verbal and nonverbal behavior. *J. Nonverbal Behav.* 24, 239–263.

Zvi, L., Nachson, I., and Elaad, E. (2012). Effects of coping and cooperative instructions on guilty and informed innocents' physiological responses to concealed information. *Int. J. Psychophysiol.* 84, 140–148.