International Journal of Psychophysiology 69 (2008) 61–68

Contents lists available at ScienceDirect

International Journal of Psychophysiology

journal homepage: www.elsevier.com/locate/ijpsycho

Electrodermal and phasic heart rate responses in the Guilty Actions Test: Comparing guilty examinees to informed and uninformed innocents ☆

Matthias Gamer [a,*], Heinz Werner Gödert [b], Alexander Keth [c], Hans-Georg Rill [b], Gerhard Vossel [b]

[a] Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Germany
[b] Department of Psychology, Interdisciplinary Research Group Forensic Psychophysiology, Johannes Gutenberg-University Mainz, Germany
[c] I. Medical Department, Division of Endocrinology and Metabolic Diseases, Johannes Gutenberg-University Mainz, Germany

ARTICLE INFO

Article history:
Received 10 July 2007
Received in revised form 11 September 2007
Accepted 3 March 2008
Available online 10 March 2008

Keywords:
Guilty Knowledge Test
Guilty Actions Test
Polygraph
Psychophysiological detection
Mock-crime
Orienting response
Habituation

ABSTRACT

The present mock-crime study concentrated on the validity of the Guilty Actions Test (GAT) and the role of the orienting response (OR) for differential autonomic responding. $N=105$ female subjects were assigned to one of three groups: a guilty group, members of which committed a mock-theft; an innocent-aware group, members of which witnessed the theft; and an innocent-unaware group. A GAT consisting of ten question sets was administered while measuring electrodermal and heart rate (HR) responses. For informed participants (guilty and innocent-aware), relevant items were accompanied by larger skin conductance responses and heart rate decelerations whereas irrelevant items elicited HR accelerations. Uninformed participants showed a non-systematic response pattern. The differential electrodermal responses of informed participants declined across the test. With respect to the HR data, however, no habituation was observed. Findings suggest that GAT results could not exclusively be interpreted by referring to the OR.

1. Introduction

The purpose of the present study was to compare phasic heart rate changes and electrodermal responses of guilty examinees, informed and uninformed innocents in the Guilty Actions Test (GAT). Moreover, the role of the orienting response (OR) for differential autonomic responding in the GAT and related techniques should be assessed. The GAT (cf. Bradley et al., 1996) is a modified version of the Guilty Knowledge Test (Lykken, 1959), a special technique designed for the detection of guilty subjects in criminal investigations. The Guilty Knowledge Test (GKT) is based on the assumption that suspects who possess knowledge about specific crime related details will be physiologically more reactive when confronted with these details than when confronted with comparable items not related to the crime (Lykken, 1959). Each crime-relevant item is presented to suspects in sets consisting of similarly plausible, but not crime related, alternatives (irrelevant items). Thus, a typical multiple-choice question in a GKT might relate to the kind of weapon used in a murder case ("Mr. X was killed with… a) a baseball bat?, b) a rifle?, c) an axe?, d) a pistol?, e) a knife?"). In the standard version of the GKT the suspect is instructed to answer "no" to each alternative. It is assumed that the weapon actually used is known only by the murderer. Therefore, this weapon has a special meaning for the guilty subject, but not for innocent suspects. Accordingly, only the guilty subject is expected to exhibit stronger autonomic responses to the relevant item than to the irrelevant items of a set whereas innocent suspects should show a non-systematic response pattern.

This hypothesis has been supported by a large body of research that mainly focused on electrodermal response differences between relevant and irrelevant items. In a recent meta-analysis, Ben-Shakhar and Elaad (2003) found high effect sizes for the differentiation of informed and uninformed subjects using electrodermal data in the GKT and related paradigms. Additionally, they identified three moderator variables. A larger number of question sets led to a higher effect size than the use of only few multiple-choice questions. Moreover, an interactive effect of the subject's motivation and the mode of responding during the test was found. Especially under low motivational conditions, a deceptive denial of the relevant item within each question set was associated with a larger effect size compared to a silent condition without an overt verbal response.

☆ The authors are indebted to Katrin Dumbert, Kirsti Maron, and Miriam Dörr for their assistance during data collection and analysis. Additionally, we thank Gershon Ben-Shakhar, Bruno Verschuere, Michael T. Bradley, and an anonymous reviewer for commenting on an earlier version of this manuscript.

* Corresponding author. Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Martinistr. 52, D-20248 Hamburg, Germany. Tel.: +49 40 42803 7160; fax: +49 40 42803 9955.

E-mail addresses: m.gamer@uke.uni-hamburg.de (M. Gamer), hwgoedert@gmx.de (H.W. Gödert).

0167-8760/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijpsycho.2008.03.001

Two critical aspects of the original GKT prompted Bradley and colleagues to slightly change the test format. First, the processes of recognition and lying are confounded, i.e., it remains unclear, to what extent increased physiological responding to the relevant items is due to recognition of crime related details and/or deceptive denial of these relevant items (see also Furedy and Ben-Shakhar, 1991). Second, the forensic application of the GKT depends on the precondition that no crime related details have been spread to the public. If this condition is not fulfilled, informed innocents could fail the GKT and as a consequence could be considered guilty. In order to solve these problems, the wording of the questions is modified in the GAT. Whereas the GKT-questions only refer to knowledge about crime details ("Mr. X was killed with…"), the GAT-questions, additionally, refer to the subject's guilt ("Did you kill Mr. X with…"). Again, all items have to be denied in order to appear innocent. Several experimental mock-crime studies found that innocents who were aware of the crime related details showed smaller response differences between relevant and irrelevant items in the GAT than did guilty subjects (Ben-Shakhar et al., 1999; Bradley et al., 1996; Bradley and Rettinger, 1992; Bradley and Warfield, 1984). Regarding the false alarms, however, a larger proportion of informed innocents (approximately 50% across studies) failed the GAT than did innocents unaware of the critical details (approximately 0% to 10% across studies).

The differential responding of guilty subjects to relevant and irrelevant items in the GKT and related techniques has been repeatedly interpreted with reference to the concept of the orienting response. The OR is a complex of behavioral and physiological reactions evoked by novel, unexpected or unpredictable stimuli (Sokolov, 1963). The OR aims at effectively preparing the organism to cope with these environmental conditions by an involuntary capture of attention and an improvement of stimulus perception. Repetition of a stimulus leads to a gradual decline of the OR magnitude, a process known as habituation. Sokolov (1963) argued that certain stimuli comprise a signal value that is capable of evoking an enhanced OR and thus preparing the organism for action. According to the notion of Lykken (1974), the relevant items embedded in the GKT-questions presentation, have this sort of significance or signal value for guilty subjects. Therefore, they evoke a stronger OR that is more resistant to habituation compared to irrelevant items (see also Ben-Shakhar and Furedy, 1990, p. 111 ff.).

More evidence for an OR theoretical account stems from studies using phasic pulse or heart rate changes as dependent variables in the GKT. Early mock-crime studies (Bradley and Ainsworth, 1984; Bradley and Janisse, 1981) as well as recent research (Gamer et al., 2006; Verschuere et al., 2004) consistently reported a relative reduction of the phasic heart rate (HR) following relevant as compared to irrelevant items. By analyzing stimulus-related HR-trends in detail, Gamer et al. (2006) found an initial HR acceleration for relevant as well as for irrelevant items in groups of guilty and innocent participants. However, only after the presentation of relevant items to guilty subjects, the phasic HR showed a marked deceleration, that peaked around 8 s after stimulus onset. The initial acceleration was interpreted as a correlate of the subject's verbal denial whereas the deceleration that most clearly distinguished between both experimental groups was thought to be related to attentional processes (see also Raskin and Hare, 1978). As the participants were requested to immediately deny each GKT question, both above mentioned processes were confounded in this study to an unknown degree. In a work by Verschuere et al. (2004), participants were not requested to respond verbally to the GKT items. In this case, the initial acceleration of the phasic HR was absent. Instead, crime related details elicited stronger HR decelerations compared to irrelevant items. This relative reduction of the HR following relevant items has been interpreted as an index of the orienting response (see also Graham and Clifton, 1966; Turpin, 1986) and thus seems to fit with the above mentioned understanding of the SCR pattern.

Taken together, the present study focused on two major issues. First, we were interested in whether the HR responses would follow the SCR pattern in a GAT examination which included a group of informed innocents. On the basis of former studies using only SCR amplitudes as dependent measure in the GAT, we expected smaller response differences between relevant and irrelevant items in the group of informed innocents as compared to guilty subjects. Second, the current study aimed at investigating the course of habituation of SCRs across the test and to relate these results to the HR responses. These analyses were performed to improve the understanding of differential physiological responding in the GAT with respect to the OR concept (Verschuere et al., 2004). According to Turpin (1986), the HR deceleration can be regarded as an index of the OR. Thus, if the physiological responding in a GAT examination mainly relies on the OR concept, the HR responses should habituate across the test. Moreover, relevant items should elicit larger responses than irrelevant items if they are recognized by the examinee (Lykken, 1974). The current study aimed at testing these predictions in a GAT examination. The basic research questions prompted us to maximize the internal validity of the current study while accepting potential losses of external or ecological validity. As will become clear in the methods section, guilty participants and informed innocents were confronted with a highly comparable experimental situation for this reason.

2. Methods

2.1. Participants

A total of 108 women participated voluntarily in the experiment in exchange for reward of at least 5 EUR (an additional amount of 7.50 EUR was paid for successfully passing the polygraph examination). They were recruited by means of flyers, placards and announcements. Most of them were students. On the arrival for the experimental session, written informed consent was obtained from all examinees. The data of three participants had to be excluded from analysis because of flawed physiological recordings. The mean age of the remaining sample ($N=105$) was 26 years ($SD=9.6$ years) with a range from 17 to 67 years.

2.2. Design and procedure

The experimental design consisted of the between-subjects factor experimental condition (guilty, innocent-aware, innocent-unaware) and the within-subject factor item type (relevant, irrelevant). All participants were randomly assigned to one of the three experimental conditions after arriving at experimenter A's (second author) office. The experimenter handed out written instructions appropriate for each individual's particular condition. Participants' questions concerning the written instructions were immediately answered by experimenter A.

Subjects in the *guilty condition* ($n=36$) were instructed to commit a mock-crime, i.e., to steal money from Professor Kunze's office (Professor Kunze was a fictive person). They were told that the money was deposited in a hidden box which was locked by a ten-digit combination lock. Since Professor Kunze was not very good at retaining digits in memory, he had written down the combination-digits on ten slips of paper, each slip containing one digit, and hidden these slips of paper at different places of his office. The slips of paper did not only contain information about the digits but also about the place where the slip of paper with the next digit of the digit combination was hidden. The ten slips of paper were located (1) in a desk drawer with a *Germany-sticker* on it, (2) under a *cactus*, (3) under a *porcelain dog*, (4) in the saddlebag of a *yellow bicycle* parked in the office, (5) behind a *picture of cows*, (6) under a *box containing water bottles*, (7) in the pocket of a *leather jacket*, (8) under a bowl containing *apples*, (9) behind a *darts board*, (10) under a *red carpet*. The slip of paper with the last digit also contained information about the place where the money box was hidden. Incidentally, the guilty subjects had come to know the place, where Professor Kunze had hidden the slip of paper containing the first digit of the digit combination. Thus, the guilty subjects just had to go to Professor Kunze's office, sequentially trace the ten slips of paper, write down the ten digits for the digit combination lock, open the lock, and take all the money deposited inside the box (50 EUR). When the subjects in the guilty condition committed the theft another woman was present, too, in Professor Kunze's office. Subjects in the guilty condition were told in advance that there would be another person in the office who allegedly took part in the experiment as a potential witness of the mock-crime. Participants in the guilty condition were instructed to try to mask the theft and to give the impression that they were innocent of the crime. Actually, the alleged witness was a confederate of the experimenter who pretended to be a naive participant. For this purpose, the confederate performed the same cleaning tasks as did the participants in the innocent-aware condition.

Subjects in the *innocent-aware condition* (n=34) were instructed to clean up Professor Kunze's office and, thereby, become witnesses of the money theft. They were told to go to Professor Kunze's office where they would find a list of ten cleaning tasks Professor Kunze asked them to dispatch. They were to (1) tighten a loose screw of the desk drawer with the *Germany-sticker* on it, (2) water the *cactus*, (3) dedust the *picture of cows*, (4) wipe the *porcelain dog*, (5) take down the darts from the *darts board*, (6) put the *leather jacket* on the hanger, (7) clean the *red carpet* with a vacuum cleaner, (8) put the *yellow bicycle* on the kickstand, (9) put the *water bottles* back into the box, (10) put the *apples* back into the bowl. Furthermore, the subjects in the innocent-aware condition were instructed that when they would be dispatching these tasks, another person would enter the office and commit a theft. In the instructions for the innocent-aware subjects the thief was alleged to be a genuine experimental participant that had been assigned to the guilty condition. In truth, the thief was a confederate of the experimenter who committed the theft in the same manner as did the actual naive subjects in the guilty condition. Thus, the subjects in the innocent-aware condition became aware of the same ten critical crime details as did the subjects in the guilty condition, first, by witnessing the theft, and second, by the critical crime details also being integral parts of the cleaning job (above, the critical details are printed in italics).

After committing or witnessing the mock theft, subjects in the guilty and innocent-aware conditions returned to the office of experimenter A in order to get further instructions. From now on, instructions were identical for all participants, i.e. for guilty subjects, innocent-aware subjects, and innocent-unaware subjects. For the subjects in the *innocent-unaware condition* (n=35), the experiment started at this point. They neither had committed the mock-theft nor had they become aware of the ten critical crime details by witnessing the mock-theft or carrying out innocent activities involving the critical details.

All subjects were instructed that they were suspected in a case of a money theft and that they would have an opportunity to demonstrate their innocence in a polygraph test examination. Further, they were told that the polygraph is very effective and that innocent subjects have a very good chance of being found innocent. If subjects were guilty, then they probably would be found guilty. As an incentive to appear innocent, all subjects were promised 7.50 EUR for successfully passing the polygraph examination. Subjects were also cautioned that because the polygraph examiner was completely blind to their guilt or innocence, they had to be cooperative and maintain their innocence, from the moment of meeting the examiner to the end of the examination. They were told that the examiner would be alert to any clues that might indicate guilt.

At the end of the examination, participants of the guilty and the innocent-aware group were asked to complete a memory test consisting of two parts. First, a cued recall test was administered without explicitly naming the alternatives. In this test, only the initial GAT questions were presented and the participants were asked to freely recall the correct item. Afterwards, a multiple choice format, consisting of all items asked during the GAT, was used as a recognition test of the respective items.

2.3. Apparatus, stimulus presentation and data collection

After being introduced to the polygraph test examiner (experimenter B) subjects sat in a comfortable semi-reclining chair and were instructed for the GAT-examination and hooked up for skin conductance and electrocardiogram measurement. The examination took place in a soundproof chamber which was dimly lit so that visual stimuli were easily visible on the display. The temperature in the chamber remained between 21 and 23°C. The GAT consisted of ten sets of items, each set containing one relevant item and five irrelevant alternatives. The relevant items were the ten critical crime details stressed earlier in the procedure section. The serial position of the relevant item varied pseudorandomly between GAT questions. The first item in each set was always an irrelevant item which served as a buffer and was not considered in later data analyses. Subjects were instructed to answer "no" to all items. Thus guilty subjects would be lying to the relevant items whereas innocent subjects would be telling the truth. A translated example of a GAT-question follows: "What colour was the carpet in the room you stole the 50 EUR from? Was it ... a) blue?, b) white?, c) yellow?, d) green?, e) red?, f) black?"

The GAT-questions and items were presented both visually and auditorily. Presentation and timing of visual and auditory stimuli were provided automatically by a specially developed computer program. For this purpose, prefabricated bitmap files and wave files were displayed simultaneously on a computer monitor and on loudspeakers. The interval between the onsets of subsequent GAT-items was 30 s. The visual presentation of each item took between 8 and 10 s. We slightly varied this interval to reduce anticipatory effects on physiological responses. The visual offset of each item served as imperative stimulus for the participants to give their verbal answer ("no").

Skin conductance was measured using a bipolar recording with two Hellige Ag–AgCl electrodes (0.8 cm diameter) filled with 0.05 M NaCl electrolyte that were connected to a constant voltage system (0.5 V). The electrodes were placed at the thenar and hypothenar eminences of the left hand. Skin conductance data were frequency-modulated by means of a transducer, registered by a digital I/O counter board and recorded at 10 Hz by a conventional personal computer outside the measurement chamber. An electrocardiogram (ECG) was obtained from two Hellige Ag–AgCl electrodes filled with electrode paste and attached to the manubrium sterni and the left lower rib cage, the reference electrode being placed on the right lower rib cage. The ECG-signal was digitised and registered by a Kölner Varioport-System polygraph with a sampling rate of 512 Hz.

2.4. Response scoring and analysis

In the analyses, we chose to concentrate on the physiological responses that were elicited by the stimulus presentation and not on those that accompanied the subjects' verbal denial. Other studies that compared electrodermal responses between these stages found a higher differentiation of relevant and irrelevant items in the first time window that comprised the stimulus presentation (e.g. Furedy and Ben-Shakhar, 1991).

Regarding the skin conductance data, we measured the artefact free amplitude of the skin conductance responses (SCR) that began between 1 and 3 s after stimulus onset as change in μS. Subsequently, the values were converted to log μS according to the formula provided by Venables and Christie (1980).

The ECG-data were first exported to ASCII-files. Afterwards, the R-waves were detected and R–R intervals in ms were converted into HR (beats per minute, bpm). Following the proposal of Velden and Wölk (1987), a second-by-second sampling of the HR was applied. This procedure implied a weighting of the HR values of each cardiac cycle depending on the relative amount of time they extended within each 1 s real time interval. The HR in the last second prior to question onset represented the prestimulus baseline. Poststimulus difference scores

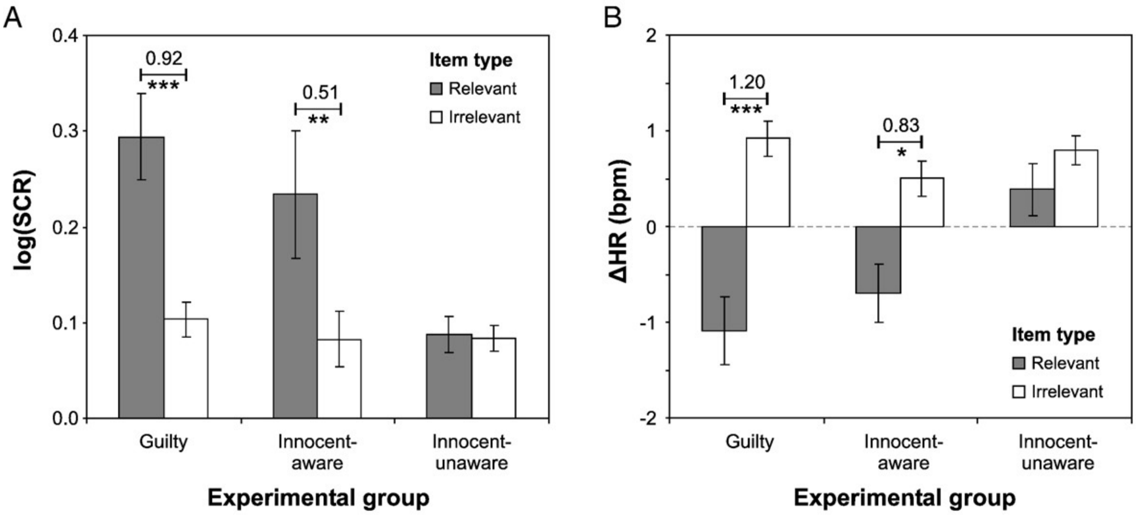

Fig. 1. Log-transformed skin conductance response (SCR) amplitudes (panel A) and mean phasic heart rate (ΔHR) in the first 8 s following item-onset (panel B) as a function of experimental group (guilty, innocent-aware, innocent-unaware) and item type (relevant, irrelevant). Differences between item types in each group were revealed by pairwise *t*-tests. Significance stars correspond to Bonferroni corrected *p*-values, * $p<.05$, ** $p<.01$, ***, $p<.001$. Cohen's *d* (Cohen, 1988, p. 19 ff.) is reported as an effect size index for all statistically significant effects. Error bars indicate standard errors of the mean.

(ΔHR) were derived by subtracting the prestimulus baseline value from the HR-score of each poststimulus-second.

Additionally, the mean HR of the first 8 poststimulus seconds of each item was computed as a more general measure of individual responsivity. The first 8 poststimulus seconds were chosen as scoring interval, since the visual presentation interval of each GAT-item, the end of which prompted the participants to give their verbal answer, also comprised at least 8 (maximally 10) poststimulus seconds (see above). Thus, the HR in the first 8 poststimulus seconds was not potentially affected by motor activity associated with speech production.

According to the suggestions by Jennings (1987), the Huynh–Feldt procedure was applied to correct for potential violations of the sphericity assumption in repeated-measures analyses of variance (ANOVAs) involving more than one degree of freedom in the enumerator. A rejection region of $p<.05$ was used for all statistical tests. For statistically significant effects, we report Cohen's *d* (for a comparison of two means, Cohen, 1988, p. 19 ff.) or Cohen's *f* (for main and interaction effects in the ANOVAs, Cohen, 1988, p. 273 ff.) as effect size estimates.

3. Results

3.1. Memory tests

To ensure, that differential physiological responses during the GAT were not due to memory differences between both informed groups, *t*-tests were carried out on the sum of correctly recalled and recognized items. Memory for relevant items did not differ significantly between guilty subjects (cued recall: $M=9.6$, $SD=1.2$; recognition: $M=9.7$, $SD=0.9$) and informed innocents (cued recall: $M=9.9$, $SD=0.3$; recognition: $M=9.9$, $SD=0.3$), neither in the cued recall, $t(68)=1.39$, $p=.17$, nor in the recognition test, $t(68)<1$.

3.2. Overall response profile

Previous studies have shown that participants recognizing the crime related item show a differential physiological response pattern to relevant and irrelevant items whereas uninformed participants respond in a non-systematic way to both of these item types (Ben-Shakhar and Elaad, 2003). In a first step, we were interested whether this overall response pattern was also obtained in the current study. Therefore, we separately averaged the responses to all relevant and all irrelevant items across the test. Regarding the HR data, we averaged the ΔHR-values for each of the first 8 s following item-onset to check for a differential temporal course of the HR responses as a function of experimental group and item type.

To statistically examine the SCR response pattern, we conducted a 3×2 repeated measures ANOVA with the group factor experimental condition (guilty, innocent-aware, innocent-unaware) and the within-subject factor item type (relevant, irrelevant) on the log-transformed SCR amplitudes. We obtained a significant condition×item type interaction, $F(2,102)=10.41$, $p<.001$, $f=0.18$, as well as a significant main effect of item type, $F(1,102)=42.51$, $p<.001$, $f=0.26$, and a marginally significant main effect of the group factor, $F(2,102)=3.02$, $p<.10$, $f=0.21$. Comparisons between the responses to both item types within each group by means of pairwise *t*-tests revealed significant differences only for informed participants (guilty and innocent-aware, see Fig. 1A). In these groups, relevant items were accompanied by larger SCR amplitudes than irrelevant items.

Regarding the HR data, we conducted a 3×2×8 repeated-measures ANOVA with the group factor experimental condition and the two within-subject factors item type and second (poststimulus seconds 1–8)[1]. The mean phasic HR-changes as a function of experimental condition and item type are depicted in Fig. 2. The ANOVA yielded a significant experimental condition×item type-interaction, $F(2,102)=5.52$, $p<.01$, $f=0.16$. Furthermore, a significant second×item type-interaction, $F(7,714)=9.73$, $\varepsilon=.61$, $p<.001$, $f=0.11$, a significant main effect of second, $F(7,714)=10.99$, $\varepsilon=.44$, $p<.001$, $f=0.16$, a significant main effect of item type, $F(1,102)=37.08$, $p<.001$, $f=0.29$, and a significant main effect of experimental condition, $F(2,102)=4.51$, $p<.05$, $f=0.15$ were obtained. Most importantly, the three way interaction of experimental condition, item type and poststimulus second was not statistically significant, $F(14,714)=1.59$, $\varepsilon=.61$, $p=.13$, $f=0.06$. Thus, the mean HR across all 8 poststimulus seconds and not the temporal trend differed as a function of experimental group and item type.

[1] We also analyzed the absolute heart rate (HR) data to check whether a differential anticipatory deceleration might have contributed to the observed response form. A 3×2×9 repeated-measures ANOVA with the group factor experimental condition and the two within-subject factors item type and second (last prestimulus and poststimulus seconds 1–8) on the absolute HR data yielded virtually identical results as the analysis of the HR changes with the last prestimulus second serving as baseline. Thus, differences in prestimulus activity did not contribute to the differential HR response to relevant and irrelevant items for participants with crime related knowledge.

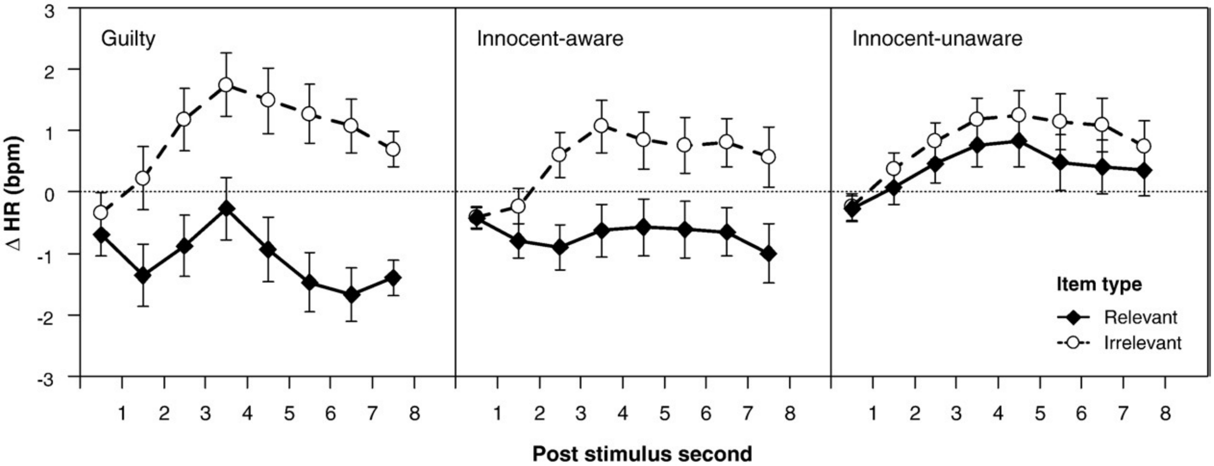

Fig. 2. Phasic heart rate (ΔHR) in the first 8 s following item-onset, as a function of experimental group (guilty, innocent-aware, innocent-unaware) and item type (relevant, irrelevant). Error bars indicate standard errors of the mean.

To further clarify the experimental condition × item type-interaction, comparisons of the mean HR-responses to both item types were accomplished within each group by means of pairwise *t*-tests. Comparable to the SCR data, significant differences between item types did only occur for informed participants (guily and innocent-aware, see Fig. 1B). In these groups, relevant items elicited a HR deceleration, whereas irrelevant items were accompanied by an accelerative HR response. Participants in the innocent-unaware condition showed a small HR acceleration irrespective of the item type[2].

3.3. Habituation of the differential responses across the test

In order to examine whether the differential responses to relevant and irrelevant items in both groups of informed participants habituated across the test, we conducted two 2×2×10 repeated measures ANOVAs with the group factor experimental condition (guilty, innocent-aware) and the within-subject factors item type (relevant, irrelevant) and multiple-choice question (factor levels 1 to 10) on the log-transformed SCR amplitudes and the average HR-responses. Regarding the SCR data, we obtained a significant interaction of item type and question, $F(9,612)=2.18$, $\varepsilon=.62$, $p<.05$, $f=0.05$. Moreover, significant main effects of item type, $F(1,68)=31.84$, $p<.001$, $f=0.21$, and question were revealed, $F(9,612)=2.57$, $\varepsilon=.53$, $p<.05$, $f=0.08$. All other effects were not significant. In contrast to these results, we did only obtain a significant main effect of item type in the analysis of the HR data, $F(1,68)=37.69$, $p<.001$, $f=0.20$. All other effects and most importantly all effects that incorporated the temporal factor of the multiple-choice question in the test were not significant. Thus, the differential HR responses to relevant and irrelevant items remained stable across the test in both groups of informed participants, whereas the SCR response differences decreased across the test in both groups. This response pattern is depicted in Fig. 3.

To examine whether changes in SCR and HR were correlated across the test, we calculated the response differences between the relevant and the four irrelevant items for each physiological measure and each multiple-choice question of the test. In a second step, we computed individual correlation coefficients between these measures across all questions for each participant. The average of within-subject correlations for both groups of informed participants was $r=.08$. Thus, temporal changes in SCR and HR were unrelated on a within-subject basis.

[2] The existence of these accelatory and decelatory HR responses to each item type in all experimental groups was confirmed by a series of *t*-tests for significant differences from 0. All HR responses differed significantly from 0, except the response to relevant items in the group of uninformed innocents. Thus, with the exception of this response, all item types elicited a substantial change of the HR in each experimental group.

3.4. Validity of the GAT

To evaluate the diagnostic utility of skin conductance and heart rate data in the GAT setting, ROC curves (see Ben-Shakhar and Elaad, 2003; National Research Council, 2003) were computed on the basis of the response differences between relevant and irrelevant items. For that purpose, each informed group (guilty and innocent-aware) was contrasted with the group of uninformed innocents. Because a negative difference between the heart rate response to relevant and irrelevant items reflects recognition of the critical item, this value was multiplied by −1 to score in the same direction as the SCR difference. The area under the ROC curve can be regarded as an index of separation between the informed and the uninformed group. The area statistic varies between 0 and 1. A value of 0.5 can be regarded as a random classification, an area statistic of 1 indicates that there is no overlap between the parameter distributions of both groups, i.e. they could be separated perfectly. Bamber (1975) described a method for estimating the variance of the area statistic, which allows for a computation of confidence intervals for the true area. Using this method, we computed 95% confidence intervals for the group comparisons. As can be seen from Table 1, guilty participants could be significantly differentiated from the innocent-unaware group using skin conductance as well as heart rate. The differentiation of informed and uninformed innocents was less clear when relying on the HR responses (note that the lower bound of the confidence interval equals 0.5 which could be regarded as random differentiation).

A statistical comparison of the area statistics of both physiological measures revealed that skin conductance responses were associated with a significantly larger area statistic than HR responses when contrasting informed innocents with the innocent-unaware group, $z=2.09$, $p<.05$, whereas the area statistics of both measures did not differ significantly when comparing guilty subjects to the innocent-unaware group, $z=1.58$, $p=.11$ (see Hanley and McNeil, 1982; Metz et al., 1984; for the calculations). A statistical comparison of the area statistics between both group contrasts (guily/innocent-unaware vs. innocent-aware/innocent-unaware) revealed no significant difference for either physiological measure. That is, on the basis of these data, it can not be reliably concluded that guilty participants showed significantly larger response differences than informed innocents using electrodermal or HR responses.

4. Discussion

In the present study, skin conductance and heart rate responses during a GAT examination were compared between three groups: A guilty group, members of which committed a mock-theft, an innocent-aware group, members of which became aware of the crime details by

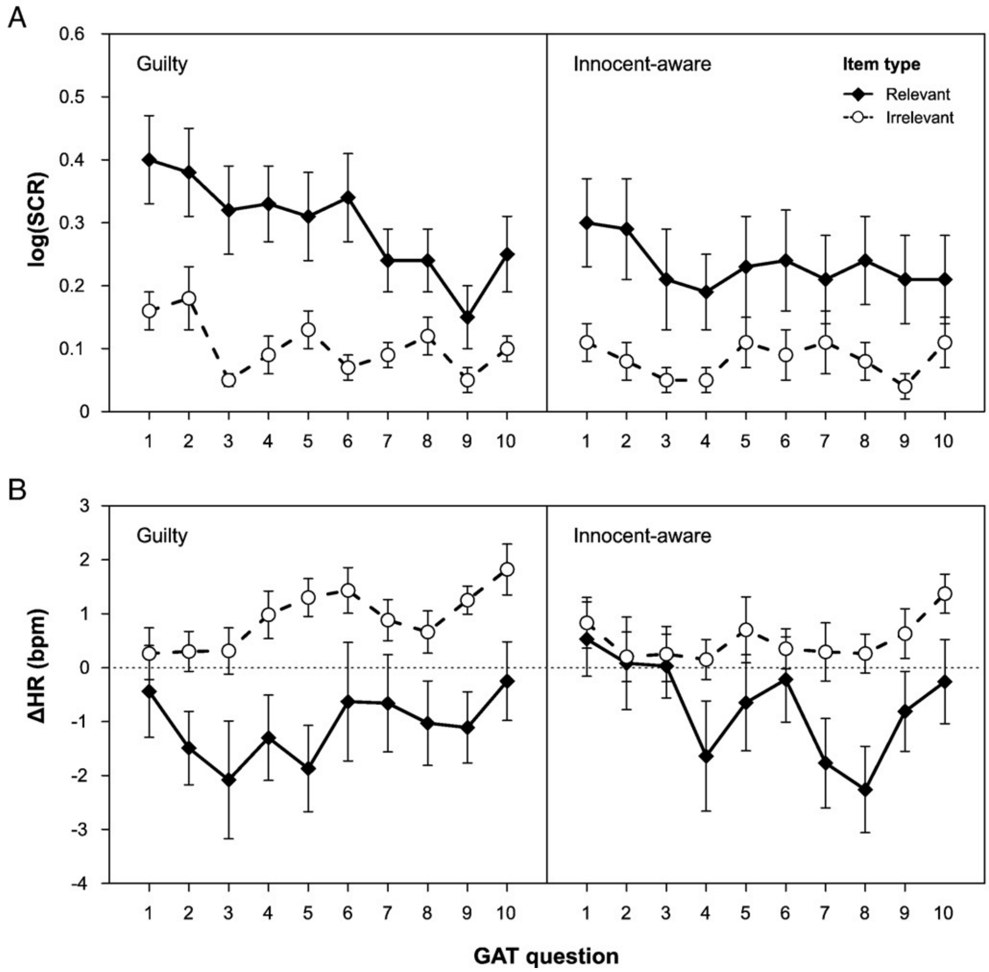

Fig. 3. Log-transformed skin conductance response (SCR) amplitudes (panel A) and mean phasic heart rate (ΔHR) in the first 8 s following item-onset (panel B) as a function of experimental group (guilty, innocent-aware), item type (relevant, irrelevant) and position of the multiple-choice question in the test. Error bars indicate standard errors of the mean.

witnessing the theft, and an innocent-unaware group, members of which neither committed nor witnessed the theft. Significant response differences between relevant and irrelevant items were found for guilty subjects as well as for informed innocents in both physiological measures. On the other hand, no systematic response differences were obtained for innocents that remained ignorant to the crime related details. This overall response pattern closely fits to other research on electrodermal (Ben-Shakhar and Elaad, 2003) and heart rate responses (Bradley and Ainsworth, 1984; Bradley and Janisse, 1981; Gamer et al., 2006; Verschuere et al., 2004) in the GKT. Although the effect sizes of the response differences between relevant and irrelevant items tended to be smaller for informed innocents as compared to guilty participants, this effect did not emerge statistically. Thus, recognition of crime related details seems to have a larger impact on the physiological response pattern than actual execution of the mock crime. This contradicts other studies on the GAT that reported significant differences between guilty participants and innocents who were aware of crime related details (Ben-Shakhar et al., 1999; Bradley et al., 1996; Bradley and Rettinger, 1992; Bradley and Warfield, 1984). Two aspects should be taken into account to explain this discrepancy: First, the GAT questions targeted 10 details that were rather peripheral for the offence because they mainly represented the room's furniture. As a consequence, it was difficult to emphasize the action facet of the offence in the question wording. We tried to account for this fact by including a self-referring accusation (e.g. "you stole") in each question but the participants might have put more emphasis on the knowledge instead of the action facet of each question. Thus, the examination might have been perceived as somewhat similar to a GKT which could have contributed to the lack of differentiation between both informed groups. If this was true, precise wording of the questions would be a crucial issue that should be systematically investigated by future research.

A second issue might be related to the stringent experimental control in the current study. Participants of the guilty as well as of the innocent-aware group were obtrusively exposed to all crime related details. Guilty subjects had to attentively go through all details to obtain the location of the money box and the code for the ten-digit combination lock. Informed innocents, on the other hand, witnessed the theft and were confronted with all crime related details by accomplishing the cleaning job. As a consequence, both groups scored nearly perfectly in the memory tests and showed a very similar response pattern in the GAT examination. Interestingly, these results resemble data from a so called "Innocent Associations Group" in the study of Bradley and Warfield (1984)[3]. Participants of this group

Table 1
Area under the ROC curves and associated 95% confidence intervals (CI) for skin conductance and heart rate data

Group comparison	Skin conductance		Heart rate	
	Area	95% CI	Area	95% CI
Guilty/Innocent-unaware	.87	(.77, .94)	.75	(.63, .85)
Innocent-aware/Innocent-unaware	.82	(.71, .90)	.64	(.50, .76)

Note. N=36, 34, and 35 in the guilty, innocent-aware, and innocent-unaware groups, respectively.

[3] We are grateful to Michael T. Bradley for raising this issue.

became aware of crime-relevant information by actively handling the items in a way that was very similar to the cleaning scenario of the current study. When comparing this group to guilty examinees and several other groups of informed innocents, it turned out that they scored higher in the GAT than innocents that were only passively exposed to crime related details. As a consequence, six of these eight participants were classified guilty. Bradley and Warfield (1984) did also measure memory for crime-relevant information and paradoxically found that the "Innocent Associations Group" had lower memory scores than guilty examinees and other groups of informed innocents.

In the current study, a much larger group of participants was examined and we obtained similar memory scores for guilty examinees and informed innocents. Moreover, both groups were indistinguishable in their physiological response profile. This suggests that the context during which crime-relevant information is acquired might be essential for the physiological responding in the GAT. When an examinee actively handles the crime related objects, she has a higher risk for showing a response pattern similar to the culprit even when being innocent of the offence itself. It might be argued that such a group of innocents is not ecologically valid (an argument that was also raised by Bradley and Warfield, 1984, p. 688). In some forensic cases, however, people might become suspects because they were at the crime site at roughly the same time as the culprit. Eventually, they also became actively engaged with items that would be used for a GAT examination. The results of the current study suggest that in such (rare) cases no GAT should be administered because of the high risk for false positive outcomes.

In line with recent studies on the GKT, we observed a parallel HR deceleration and SCR increase when the examinee recognized the relevant item (Gamer et al., 2006; Verschuere et al., 2004). This pattern is especially interesting because rapid HR decelerations are primarily mediated by the parasympathetic nervous system (PNS) whereas the skin conductance is solely modulated by the sympathetic nervous system (SNS; Wallin, 1981). Thus, the complex of skin conductance increase and HR deceleration might be the result of a simultaneous or at least closely proximate coactivation of both divisions of the autonomic nervous system in the respective experimental conditions (Furedy, 1985). Comparable results were also obtained in the differentiation-of-deception paradigm (Heslegrave, 1982, see also Ben-Shakhar & Furedy, 1990, p. 142). This rather independent regulation of the PNS and the SNS instead of a single dimensional continuum extending from vagal to sympathetic dominance has been hypothesized before and corroborated by experimental data (see Berntson et al., 1991, for an overview). The experimental procedure of the GKT and related techniques obviously seems to be a candidate for provoking a coactivation of the PNS and the SNS. However, it has still to be examined by future research whether both divisions of the autonomic nervous system might respond differentially to other experimental variations of this paradigm, i.e. motivational incentives, different response modes, etc.

With regard to the second research question, we examined the course of habituation of the differential response pattern to relevant and irrelevant items in both groups of informed participants. For the electrodermal data, the response difference between relevant and irrelevant items gradually declined across the 10 GAT questions that were used in the examination. By contrast, the HR responses in both groups of informed participants showed no evidence for such a habituation across the test. This missing habituation has been observed earlier in simple OR habituation paradigms (e.g. Barry, 1983; Vossel and Zimmer, 1989; for an overview see Barry, 1996) and might indicate that HR deceleration, in contrast to SCR magnitude (Ben-Shakhar, 1977; Ben-Shakhar et al., 1975), could not simply be regarded as an index of the OR (Barry and Maltzman, 1985). This claim is further substantiated by the missing within-subject correlation between these measures that was observed in the current study. Barry considered the initial HR deceleration as a preliminary process of stimulus registration. According to this view, the HR responses to relevant and irrelevant items would not be expected to differ. We did, however, provide evidence for a differential HR responding to these item types in both groups of informed participants. The ECR1 component (first evoked cardiac response) according to Barry is usually confined to the first 2 s after stimulus onset. As we considered an 8-s interval for our analyses, our results might reflect a different underlying mechanism. The detailed trends of the HR changes elicited by the stimulus presentation (see Fig. 2) do however reveal that differential responses to both item types began very early in the current study, approximately after the first poststimulus second[4]. Verschuere et al. (2004) reported a very similar response differentiation soon after the stimulus onset. Thus, simple OR theories (e.g. Barry, 1996; Graham and Clifton, 1966) do not fully account for the response pattern that was observed in the current study.

The revised baroreceptor hypothesis by Wölk and Velden (1987, 1989; see also Wölk et al., 1989) seems to provide a plausible interpretation of our findings. According to this hypothesis, that is based on earlier work by Lacey and Lacey (1970), phasic HR-changes that are primarily modulated by the PNS, influence the cortex via the carotid baroreceptors and the thalamus, resulting in changes in sensory and sensory-motor performance. HR decelerations should exert a disinhibitory effect on the cortex and this effect should consequently facilitate a more elaborated processing of the relevant GAT-items for guilty and innocent-aware subjects. Thus, the decelerative HR responses that are triggered by relevant GAT-items may be a correlate of attentional redirection towards the current (and following) GAT-items or towards the own bodily responses with the potential aim of monitoring and controlling them. On the other hand, HR-accelerations exerting an inhibitory effect should be the consequence of unrecognized items that do not require a comparable elaborated processing. This broader theoretical framework that focuses on information processing demands and their physiological basis fits to former GKT studies demonstrating that the recognition of crime details by guilty subjects is associated with a relative decline in phasic pulse or heart rate (Bradley and Ainsworth, 1984; Bradley and Janisse, 1981; Gamer et al., 2006; Verschuere et al., 2004). Moreover, as the revised baroreceptor hypothesis makes no assumptions about a systematic decline in response strength as a function of repeated stimulus presentations, it also accounts for the current data. Thus, on the one hand, stable HR differences between relevant and irrelevant items in the GKT and related techniques could be integrated in this theoretical framework. On the other hand, it must be emphasized that the baroreceptor hypothesis lacks empirical substantiation in general as for example the expected correlation between reaction time and HR could not be detected on a within-subject basis (Iacono and Lykken, 1978). Taken together, it remains a challenging task for future research to develop an adequate theoretical framework for the GKT and related techniques. Such an account must probably widen the focus from the rather narrow OR perspective to a broader theory of information processing, memory and allocation of attentional resources to fully account for differential physiological responding in the detection of concealed information.

References

Barry, R.J., 1983. Primary bradycardia and the evoked cardiac response in the OR context. Physiol. Psychol. 11, 135–140.

Barry, R.J., 1996. Preliminary process theory: towards an integrated account of the psychophysiology of cognitive processes. Acta Neurobiol. Exp. 56, 469–484.

Barry, R.J., Maltzman, I., 1985. Heart rate deceleration is not an orienting reflex; heart rate acceleration is not a defensive reflex. Pavlovian J. Biol. Sci. 20, 15–28.

Bamber, D., 1975. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. J. Math. Psychol. 12, 387–415.

[4] In fact, a reduced 3 (experimental condition)×2 (item type)×2 (second) repeated-measures ANOVA of the HR data that was confined to the first two seconds after item onset, revealed a significant interaction of all three factors, $F(2,102)=3.36$, $p<.05$, $f=0.06$, indicating that differential responses to relevant and irrelevant items in both groups of informed subjects occurred very soon after item onset. Thus, we did indeed observe a differential ECR1 component for relevant and irrelevant items when the participant was able to distinguish between them. No such difference could be observed for innocents unaware of crime related details.

Ben-Shakhar, G., 1977. A further study of dichotomization theory in detection of information. Psychophysiology 14, 408–413.
Ben-Shakhar, G., 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., Furedy, J.J., 1990. Theories and applications in the Detection of Deception: A Psychophysiological and International Perspective. Springer, New York.
Ben-Shakhar, G., Lieblich, I., Kugelmass, S., 1975. Detection of information and GSR habituation: an attempt to derive detection efficiency from two habituation curves. Psychophysiology 12, 283–288.
Ben-Shakhar, G., Gronau, N., 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.
Berntson, G.G., Cacioppo, J.T., Quigley, K.S., 1991. Autonomic determinism: the modes of autonomic control, the doctrine of autonomic space, and the laws of autonomic constraints. Psychol. Rev. 98, 459–487.
Bradley, M.T., Ainsworth, D., 1984. Alcohol and the psychophysiological detection of deception. Psychophysiology 21, 63–71.
Bradley, M.T., Janisse, M.P., 1981. Accuracy demonstrations, threat, and the detection of deception: cardiovascular, electrodermal, and pupillary measures. Psychophysiology 18, 307–315.
Bradley, M.T., Rettinger, J., 1992. Awareness of crime-relevant information and the guilty knowledge test. J. Appl. Psychol. 77, 55–59.
Bradley, M.T., Warfield, J.F., 1984. Innocence, information, and the guilty knowledge test in the detection of deception. Psychophysiology 21, 683–689.
Bradley, M.T., MacLaren, V.V., Carle, S.B., 1996. Deception and nondeception in guilty knowledge and guilty actions polygraph tests. J. Appl. Psychol. 81, 153–160.
Cohen, J.D., 1988. Statistical Power Analysis for the Behavioral Sciences, 2nd edn. Lawrence Erlbaum, Hillsdale.
Furedy, J.J., 1985. Joint use of heart-rate and T-wave amplitude as non-invasive cardiac performance measures: a psychophysiological perspective. In: Orlebeke, J.F., Mulder, G., van Doornen, L.J.P. (Eds.), Psychophysiology of Cardiovascular Control. Plenum, New York, pp. 237–256.
Furedy, J.J., Ben-Shakhar, G., 1991. The roles of deception, intention to deceive, and motivation to avoid detection in the psychophysiological detection of guilty knowledge. Psychophysiology 28, 163–171.
Gamer, M., Rill, H.-G., Vossel, G., Gödert, H.W., 2006. Psychophysiological and vocal measures in the detection of guilty knowledge. Int. J. Psychophysiol. 60, 76–87.
Graham, F.K., Clifton, R.K., 1966. Heart-rate change as a component of the orienting response. Psychol. Bull. 65, 305–320.
Hanley, J.A., McNeil, B.J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143, 29–36.
Heslegrave, R.J., 1982. An examination of the psychological mechanisms underlying deception. Psychophysiology 19, 323 (Abstract).
Iacono, W.G., Lykken, D.T., 1978. Within-subject covariation of reaction time and foreperiod cardiac deceleration: effects of respiration and imperative stimulus intensity. Biol. Psychol. 7, 287–302.
Jennings, J.R., 1987. Editorial policy on analyses of variance with repeated measures. Psychophysiology 24, 474–478.
Lacey, J.I., Lacey, B.C., 1970. Some autonomic-central nervous system interrelationships. In: Black, P. (Ed.), Physiological Correlates of Emotion. Academic Press, New York, pp. 205–227.
Lykken, D.T., 1959. The GSR in the detection of guilt. J. Appl. Psychol. 43, 385–388.
Lykken, D.T., 1974. Psychology and the lie detector industry. Am. Psychol. 29, 725–739.
Metz, C.E., Wang, P.-L., Kronman, H.B., 1984. A new approach for testing the significance of differences between ROC curves measured from correlated data. In: Deconinck, F. (Ed.), Information Processing in Medical Imaging. Martinus Nijhoff, The Hague, pp. 432–445.
National Research Council, 2003. The polygraph and lie detection. Committee to review the scientific evidence on the polygraph. Division of Behavioral and Social Sciences and Education. The National Academies Press, Washington, DC.
Raskin, D.C., Hare, R.D., 1978. Psychophathy and the detection of deception in a prison population. Psychophysiology 15, 126–136.
Sokolov, E.N., 1963. Perception and the Conditioned Reflex. Pergamon Press, Oxford.
Turpin, G., 1986. Effects of stimulus intensity on autonomic responding: the problem of differentiating orienting and defense reflexes. Psychophysiology 23, 1–14.
Velden, M., Wölk, C., 1987. Depicting cardiac activity over real time: a proposal for standardization. J. Psychophysiol. 1, 173–175.
Venables, P.H., Christie, M.J., 1980. Electrodermal activity. In: Martin, I., Venables, P.H. (Eds.), Techniques in Psychophysiology. Wiley, Chichester, pp. 3–67.
Verschuere, B., Crombez, G., De Clercq, A., Koster, E.H., 2004. Autonomic and behavioral responding to concealed information: differentiating orienting and defensive responses. Psychophysiology 41, 461–466.
Vossel, G., Zimmer, H., 1989. Heart rate as an index of the orienting response? J. Psychophysiol. 3, 111–124.
Wallin, B.G., 1981. Sympathetic nerve activity underlying electrodermal and cardiovascular reactions in man. Psychophysiology 18, 470–476.
Wölk, C., Velden, M., 1987. Detection variability within the cardiac cycle: toward a revision of the 'baroreceptor hypothesis'. J. Psychophysiol. 1, 61–65.
Wölk, C., Velden, M., 1989. Revision of the baroreceptor hypothesis on the basis of a new cardiac cycle effect. In: Bond, N.W., Siddle, D.A.T. (Eds.), Psychobiology: Issues and Applications. Elsevier, Amsterdam, pp. 371–379.
Wölk, C., Velden, M., Zimmermann, U., Krug, S., 1989. The interrelation between phasic blood pressure and heart rate changes in the context of the 'baroreceptor hypothesis'. J. Psychophysiol. 3, 397–402.