Behavior Research Methods, Instruments, & Computers
2000, 32 (1), 111-119

Resampling approach to statistical inference: Bootstrapping from event-related potentials data

FRANCESCO DI NOCERA and FABIO FERLAZZO
University of Rome "La Sapienza," Rome, Italy

We propose the use of the bootstrap resampling technique as a tool to assess the within-subject reliability of experimental modulation effects on event-related potentials (ERPs). The assessment of the within-subject reliability is relevant in all those cases when the subject score is obtained by some estimation procedure, such as averaging. In these cases, possible deviations from the assumptions on which the estimation procedure relies may lead to severely biased results and, consequently, to incorrect functional inferences. In this study, we applied bootstrap analysis to data from an experiment aimed at investigating the relationship between ERPs and memory processes. ERPs were recorded from two groups of subjects engaged in a recognition memory task. During the study phase, subjects in Group A were required to make an orthographic judgment on 160 visually presented words, whereas subjects in Group B were only required to pay attention to the words. During the test phase all subjects were presented with the 160 previously studied words along with 160 new words and were required to decide whether the current word was "old" or "new." To assess the effect of word imagery value, half of the words had a high imagery value and half a low imagery value. Analyses of variance performed on ERPs showed that an imagery-induced modulation of the old/new effect was evident only for subjects who were not engaged in the orthographic task during the study phase. This result supports the hypothesis that this modulation is due to some aspect of the recognition memory process and not to the stimulus encoding operations that occur during the recognition memory task. However, bootstrap analysis on the same data showed that the old/new effect on ERPs was not reliable for all the subjects. This result suggests that only a cautious inference can be made from these data.

Functional inferences from studies in cognitive neuroscience are often based on sets of methodological assumptions whose validity is not always evident. For example, since in event-related potential (ERP) research the low signal-to-noise ratio makes it generally impossible to analyze the electrocerebral response to a single stimulus, such a brain response needs to be estimated, usually through an averaging procedure. The mathematical model that underlies averaging (De Weerd, 1981; De Weerd & Martens, 1978; Woody, 1967) holds that the electrocerebral activity recorded upon the presentation of a stimulus is the sum of the brain response to that stimulus (signal) and the spontaneous brain activity independent of it (noise). Furthermore, the model holds that the brain response is invariant with respect to different occurrences of the same stimulus, and that the spontaneous brain activity is drawn from a zero-mean random process. Under these assumptions, averaging gives a reliable estimate of the true brain response when applied to a large number of occurrences of the same stimulus. The assumptions on which the averaging procedure is founded, however, are often questionable. In particular, the variability of late (cognitive) potentials and the correlation between spontaneous EEG and brain response to the stimulus represent a violation of those assumptions. This violation may severely affect the reliability of ERP estimates and, consequently, the reliability of the experimental results and their interpretation. In fact, a difference between ERPs recorded in two conditions from 1 subject can be ascribed to the independent variable only if the ERP estimates were reliable (reliable within-subject effect). However, if the ERPs estimates were unreliable, the between-conditions difference should be ascribed to chance (unreliable within-subject effect).

Of course, since *across-subjects* statistical data analyses are usually performed on dependent variables (amplitude, latency, etc.) that are computed on averaged ERPs, unreliable within-subject effects of the independent variable could strongly affect the overall results. This is because an unreliable difference between ERPs in 1 subject has the same weight on the across-subjects analysis of a reliable difference in another subject. Furthermore, since in psychophysiological research small samples are often used, even a small number of unreliable differences may have large effects on the overall analysis.

One approach to the evaluation of the reliability of an observed experimental effect is provided by the bootstrap technique (Efron, 1979; Efron & Tibshirani, 1993). Resampling techniques such as the bootstrap or permutation test have already been proposed and used in psychophysiology for overall significance testing and for

Correspondence should be addressed to F. Di Nocera, Department of Psychology, University of Rome "La Sapienza," Via dei Marsi, 78, 00185 Rome, Italy (e-mail: dinocera@uniroma1.it).

Copyright 2000 Psychonomic Society, Inc.

determining the reliability of maps of brain activity across groups or across tasks within the same subject (e.g., Blair & Karniski, 1993; Di Nocera, Ferlazzo, & Gentilomo, 1996; Farwell & Donchin, 1991; Humphrey & Kramer, 1994; Karniski, Blair, & Snider, 1994; Wasserman & Bockenholt, 1989). However, except for some preliminary studies of ours (Di Nocera & Ferlazzo, 1999; Ferlazzo & Di Nocera, 1998), the bootstrap procedure has never been used in psychophysiology to test hypotheses regarding the within-subject reliability of a specific experimental effect—that is, the extent to which an effect is reliable within a particular subject and not merely in respect to the whole sample. Yet such an application of bootstrapping could be of capital importance in psychophysiology, where measures are not directly observed (they are estimated), and hence interpretation of results depends on their reliability.

Bootstrap Procedure

The bootstrap (Efron, 1979) is a computationally intensive statistical tool designed to assess statistical accuracy. An important aspect of this technique is that it sets researchers free from making unverifiable and most likely invalid assumptions about their data (e.g., probability distribution) prior to analysis. The usefulness of this approach is particularly evident when researchers are managing small samples—that is, when it is more difficult to know how accurate an estimate is.

The basic idea underlying the bootstrap is to produce a random sample (called the *bootstrap sample*), which is obtained by sampling, with replacement, from the original pool of data. The bootstrap sample is then used to compute the estimate of the parameter the researcher is interested in, and this procedure (extraction of the random sample and computation of the estimate) is repeated many times in order to create an empirical distribution of the statistic. Such a distribution usually represents a good approximation of the true (and unknown) probability distribution underlying that statistic.

In more technical terms, the bootstrap represents a general method to estimate how well a statistic $\emptyset'$ estimates a parameter $\emptyset$ when analytical procedures are not available. In other words, it estimates the standard error of $\emptyset'$.

The bootstrap procedure can be also used in hypothesis testing. For example, consider two unknown probability functions F and G. To test the null hypothesis $\mu(F) = \mu(G)$, we define $\Theta' = z - y$, where z and y are, respectively, the means of two samples drawn from F and G. The bootstrap procedure puts together all the observations from the two samples and extracts randomly, with replacement, two new bootstrap samples from the whole pool of data. This extraction procedure is repeated many times, and each time the corresponding Θ'^* is computed. Here the asterisk denotes the statistic computed on the bootstrap samples. Since the pairs of bootstrap samples are formed on a random basis, the expected value of Θ'^* is null, and the distribution of Θ'^* represents its empirical distribution under the null hypothesis. Hence, through this empirical distribution, it is possible to estimate the probability of occurrence of the observed Θ' under the null hypothesis—that is, how many Θ'^*s are equal or larger than Θ' by chance—and therefore its level of significance.

The application to ERPs studies is straightforward. The bootstrap procedure can be applied within each subject to test the hypothesis that the difference between ERPs recorded in two different conditions is due to the independent variable or, in other words, whether it is reliable or not. To do that, single-trial ERPs are first pooled together to form a unique pool of data, and then they are drawn randomly and with replacement to form two random sets of data with an equal number of single-trial ERPs. It should be noted that because of the random assignment, each bootstrap sample is not necessarily composed of an equal number of single-trial ERPs from each "real" condition. Averaging is then performed for each bootstrap sample, and the difference between the dependent variable (e.g., mean amplitude or peak-to-peak amplitude) of the two averaged ERPs is computed. Because of the random assignment to the two sets, the expected value of this difference is null. The probability distribution under the null hypothesis that no difference exists between conditions is created by repeating these two steps many times. This empirical distribution can be used to estimate the probability that the observed difference between ERPs is due to chance: If the probability is lower than the usual significance level of .05, we can reject the null hypothesis and accept that the difference is due to the experimental variable.

Through this procedure, information about the reliability of ERP estimates can be gained and used to modulate the interpretation of the results of conventional statistical analyses performed on ERP estimates: If a large number of subjects show an unreliable difference between conditions, of course, no confidence should be placed in the results of across-subjects statistical analyses, even if significant.

In order to evaluate the usefulness of the bootstrap approach to ERP reliability issues, we applied this procedure to further analyze data from a study aimed at investigating the imagery-induced modulation of ERPs to old and new words (old/new effect) recorded during a recognition memory task.

An Application of the Bootstrap Approach

In memory tasks, ERPs have been reported to differ according to whether the stimuli evoking them had been previously presented or not (for a review, see Rugg, 1995). This old/new effect has been widely studied during the last decade and refers to the larger positivity from about 400 msec onward shown by the ERPs to old items relative to the ERPs to new items. The effect has been shown using different paradigms (study–test, continuous recognition) as well as different materials (words, pictures,

notes, numbers, musical notes, but not geometrical figures (Beisteiner, Huter, Edward, & Koch, 1997) in both visual and auditory modalities (Bentin, 1987; Donaldson & Rugg, 1998; Ferlazzo, Conte, & Gentilomo, 1993a, 1993b; Ferlazzo, Di Nocera, & Di Segni, 1998; Friedman, 1990; Friedman & Sutton, 1987; Johnson, Kreiter, Russo, & Zhu, 1998; Karis, Fabiani, & Donchin, 1984; Neville, Kutas, & Schmidt, 1982; Noldy-Cullum & Stelmack, 1987; Pratt, Erez, & Geva, 1994; Rugg, 1985; Rugg, Furda, & Lorist, 1988; Rugg, Mark, et al., 1998; Rugg & Nagy, 1987; Rugg, Schloerscheidt, & Mark, 1998; Sanquist, Rohrbaugh, Syndulko, & Lindsley, 1980; Schloerscheidt & Rugg, 1997; Smith & Guster, 1993). Many of these authors have linked this effect to memory processes, but its exact meaning is still uncertain. Furthermore, the variability of the results reported in the literature suggests that this effect is not as reliable as others. In this study, we aimed at investigating the modulation of the old/new effect by the imagery value of words used as stimuli in a recognition memory task. This choice underlines the need to provide a strongly controlled experimental design to allow comparison of conventional analyses and bootstrap results.

The enhanced memory for high-imagery (HI) value words relative to low-imagery (LI) value words (Paivio, 1965) is a well-known effect on memory performance. Ferlazzo et al. (1993a) suggested that the imagery value can modulate the old/new effect on ERPs. Particularly, their results showed that a larger difference between old and new words was associated with the LI condition, supporting the hypothesis that the old/new effect depends on the recognition memory process, since nonspecific factors such as target detection operations cannot account for it. However, this interpretation may be undermined by the fact that imagery value is an intrinsic attribute of each word. With that in mind, since in a study–test paradigm the stimulus encoding also occurs during the test phase, the imagery-induced modulation might depend on this encoding process.

We should be able to verify this last hypothesis by selectively interfering with the semantic processing of the word during the study phase. In fact, if the modulation effect still emerges in such a condition, an encoding process is probably involved. If no modulation emerges in the interference condition, a memory recognition process is probably involved in the genesis of the old/new effect.

METHOD

Subjects

Seventeen subjects (5 males and 12 females) participated in this experiment. Their mean age was 25.9 years. All subjects were naive with respect to the experimental procedures.

Stimuli

One hundred sixty HI words and 160 LI words were selected from the Bartolini, Tavaglini, and Zampolli (1971) Italian norms of frequency of usage. Words were rated as HI or LI by three independent raters. Eighty HI and 80 LI words were selected from the whole set to form the study list (old words). The whole set of 320 words served as test list in the recognition memory phase of the experiment. The test list was composed of four blocks of 80 words each: HI and LI old words and HI and LI new words. Stimuli were randomly ordered in each list, and their frequency of usage was balanced across all blocks.

Procedure

Following electrode application, subjects were seated in an electrically shielded and sound-attenuated room and completed an intentional memory task in two phases. During the study phase, the 160 words from the study list (randomly ordered) were visually presented at a rate of about 1.5 sec (mean interstimulus interval [ISI]). All subjects were required to pay close attention to the stimuli because in a successive test phase they would be required to recognize them from an equal number of new words. During the study phase, 9 subjects out of 17 (Group A) were also engaged in an interfering orthographic task on the same words. They were required to press one of two buttons according to whether the current word included at least one letter *O* or not. This task should interfere with the semantic encoding of the words on which the imagery effect is based. The other 8 subjects (Group B) were not engaged in any task during the study phase.

The recognition memory test followed the study phase after a 10-min resting period. The test list was presented in four runs of 80 words each, randomly ordered across subjects. During the test phase, words were presented at a rate of 3 sec (mean ISI). Each run was followed by a 5-min rest period. During the test phase, all subjects were required to press one of two buttons as fast as possible according to whether the current word was old (previously presented) or new.

EEG Recording

The electroencephalogram (EEG) was recorded during the test session of the experiment using Beckman Ag/AgCl electrodes placed at Fz, Cz, and Pz sites (10–20 International System; Jasper, 1958) and referred to linked mastoids. Electrode impedance did not exceed 5 kΩ. The EEG was amplified with the low filter set at 35 Hz and time constant set at 1 sec.

Conventional Analyses

EEG recorded during the test session was sampled at 128 Hz for 1,000 msec beginning 150 msec prior to each word onset and averaged separately for each stimulus category (HI old words, LI old words, HI new words, and LI new words) and electrode lead (Fz, Cz, and Pz). EEG epochs were visually inspected and those containing eye movement artifacts were discarded. Trials where subjects gave the incorrect response or no response within 2 sec were not included in the averaged ERPs or any further analysis.

A four-way mixed analysis of variance (ANOVA) was performed on ERP mean amplitudes in the 400–800 msec latency range. Factors were Group (A vs. B), electrode lead (Fz vs. Cz vs. Pz), stimulus (old vs. new), and imagery (HI vs. LI). Three-way and four-

Table 1
Percentages and Standard Deviations of Hits and Correct Rejections on the Recognition Memory Test for High- and Low-Imagery Words Shown by Subjects Who Completed the Study Phase With and Without the Interference Task

	New Words		Old Words	
Imagery	%	*SD*	%	*SD*
	With Interference			
High	66.39	17.27	54.86	14.41
Low	64.44	14.31	55.28	19.37
	Without Interference			
High	56.25	30.19	54.53	17.64
Low	60.16	27.57	44.69	12.81

Table 2
Mean Reaction Times (in Milliseconds) and Standard Deviations of Correct and Incorrect Responses to High- and Low-Imagery Words Shown by Subjects Who Completed the Study Phase With and Without the Interference Task

	With Interference				Without Interference			
	New Words		Old Words		New Words		Old Words	
Imagery	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Correct Responses								
High	1,098.06	188.37	1,052.14	123.92	1,191.97	179.55	1,115.96	169.89
Low	1,164.34	236.87	1,071.22	181.46	1,229.35	214.71	1,141.79	167.11
Incorrect Responses								
High	1,112.79	196.42	1,157.75	200.44	1,141.06	152.14	1,130.23	216.55
Low	1,075.60	200.44	1,133.15	205.84	1,178.55	153.11	1,203.15	179.55

way ANOVAs were performed on behavioral data with factors including group (A vs. B), stimulus (old vs. new), imagery (HI vs. LI), and response (correct vs. incorrect, only for reaction time [RT] data). Behavioral measures were the number of hits (old words correctly recognized as old) and correct rejections (new words correctly recognized as new) and the corresponding RTs. The Geisser–Greenhouse conservative F test was used when necessary.

Bootstrap Analyses

The bootstrap technique was applied separately for each subject and electrode lead to test three different hypotheses arising from the experimental design: (1) The difference between ERPs to old and new words (independent from their imagery value) was due to chance; (2) the difference between ERPs to HI old and new words was due to chance; and (3) the difference between ERPs to LI old and new words was due to chance.

To test each hypothesis, 1,000 random resamplings from the appropriate pool of EEG epochs were conducted as described above. Each bootstrap sample was used to form two sets of randomly assigned EEG epochs. ERPs were estimated for each set, and the difference between their mean amplitudes in the 400–800 msec latency range was computed. Because of the random assignment, the

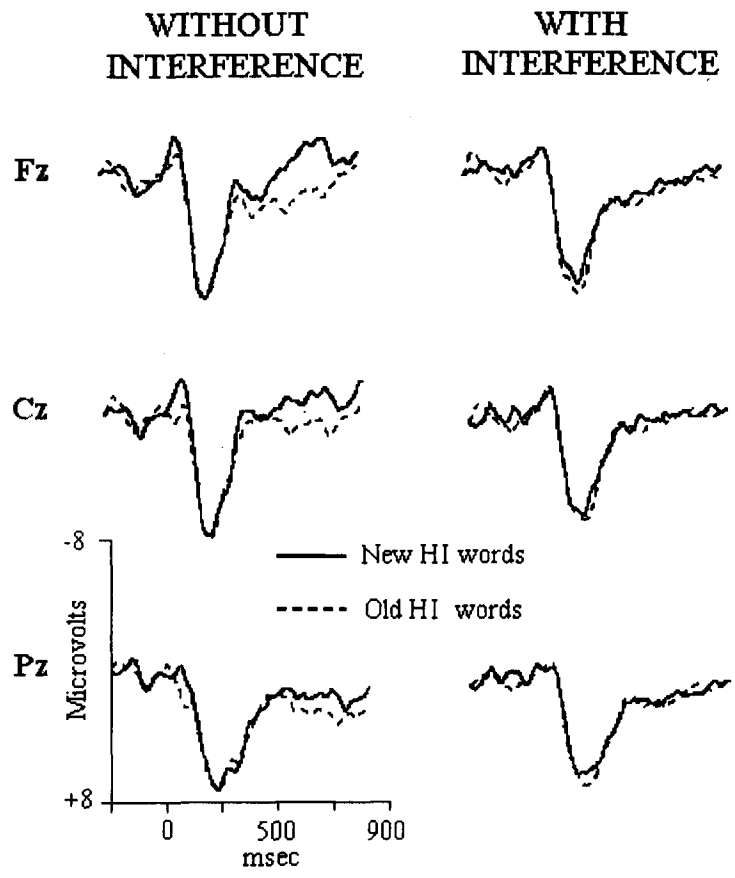

Figure 1. Grand averaged event-related potentials to high-imagery (HI) new and old words recorded in subjects who completed the study phase without and with the interference task. Stimulus onset at time 0. Negativity upward.

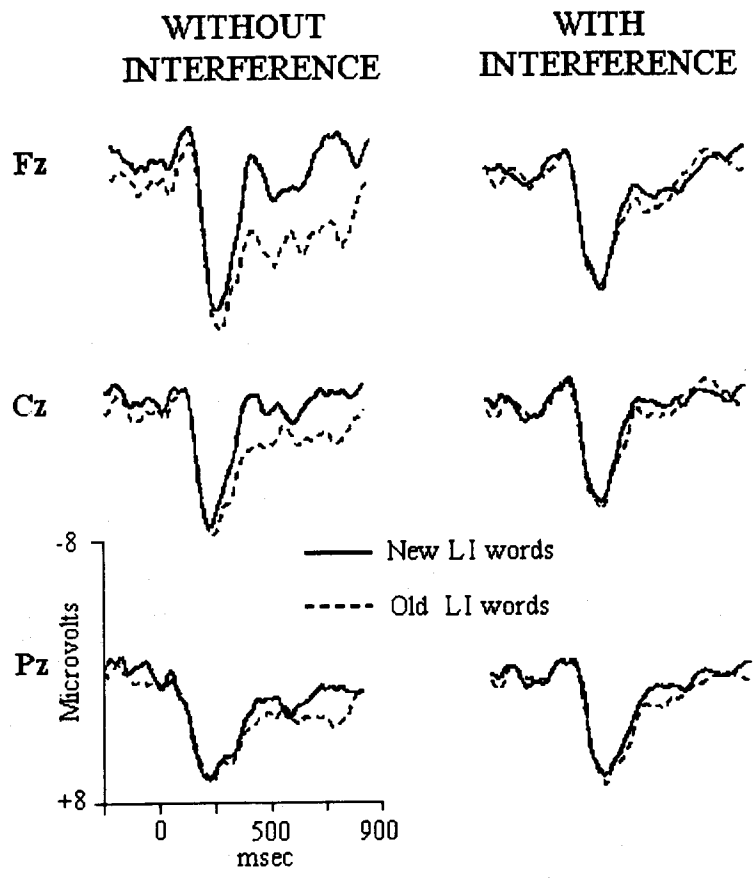

Figure 2. Grand averaged event-related potentials to low-imagery (LI) new and old words recorded in subjects who completed the study phase without and with the interference task. Stimulus onset at time 0. Negativity upward.

two sets had equal expected values; that is, the expected difference between the ERPs computed from the two sets was null. In other words, the distribution of the 1,000 bootstrap differences is the empirical probability distribution under the null hypothesis that no difference exists between ERPs. Such a distribution was used to estimate the probability associated to the observed (real) difference between ERPs: As usual, if the probability that a difference equal to or greater than the observed one was less than .05, the null hypothesis was rejected and the difference was said to be reliable. Otherwise, the observed difference could be due to chance and it was said to be unreliable.

To test the first hypothesis, ERPs to all the words were considered independently from their imagery value; to test the second hypothesis, only ERPs to the HI words were considered; to test the

Table 3
Mean Amplitudes (in Microvolts) in the 400–800 Msec Latency Range and Standard Deviations of Event-Related Potentials to New and Old Words by Imagery (High and Low), Leads (Fz, Cz, and Pz), and Group (With and Without Interference Task)

	With Interference				Without Interference			
	New Words		Old Words		New Words		Old Words	
Imagery	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
				Fz				
High	0.57	2.54	0.67	2.19	0.66	2.26	1.29	2.06
Low	0.44	1.53	0.51	1.63	0.00	2.95	3.96	3.98
				Cz				
High	−0.11	1.86	−0.03	1.56	0.52	2.18	0.88	2.59
Low	0.44	1.53	0.01	1.86	−0.23	2.89	2.66	3.40
				Pz				
High	0.69	1.32	0.71	0.87	2.44	1.42	2.69	3.28
Low	0.54	1.06	0.98	1.72	1.63	2.54	3.40	3.16

Table 4
Number of Subjects Showing a Reliable Old/New Effect by Group (With and Without Interference) and Lead (Fz, Cz, and Pz)

	With Interference			Without Interference		
Words	Fz	Cz	Pz	Fz	Cz	Pz
Old/new			3		3	4
High-imagery old/new			1	1	1	4
Low-imagery old/new		1	4		6	6

third hypothesis, only ERPs to the LI words were considered. Bootstrap analyses were performed on the same ERP trials as those used in the conventional analyses.

RESULTS

Performance

Analysis of hits and correct rejections showed only a group × stimuli × imagery interaction [$F(1,15) = 4.902$, $p = .0427$]. The effect of imagery was found only for old words and only for subjects who completed the study phase without interference [$F(1,7) = 17.821, p = .004$]. These subjects correctly recognized a significantly larger number of HI old words (54.53%) than LI old words (44.69%; Table 1).

Reaction Times

Analysis of RTs showed only a main effect of stimuli [$F(1,15) = 5.101, p = .0392$]. RTs to the old words were slightly faster (M = 1,125.67 msec, SD = 183.26 msec) than RTs to the new words (M = 1,148.96 msec, SD = 189.13 msec). No significant effect of imagery, response, or group was found on RTs (Table 2).

ERPs

For each experimental condition, averaging was computed on about 30 trials. Grand averaged ERPs to old and new words for each electrode site, imagery, and group are reported in Figures 1 and 2. Analysis of ERP mean amplitudes in the 400–800 msec latency range showed main effects of stimuli and electrode [$F(1,15) = 5.148$, $p = .0385$ and $F(2,30) = 4.544, p = .0189$, respectively], and both stimuli × imagery and stimuli × imagery × group interactions [$F(1,15) = 5.377$, $p = .0349$ and $F(1,15) = 4.253$, $p = .05$, respectively]. Simple effects showed a significant difference between ERPs to old and new words only for LI items and only for subjects who completed the study phase without interference [$F(1,8) = 10.372, p = .015$]. The difference between ERPs to old and new words was considerably larger for LI words

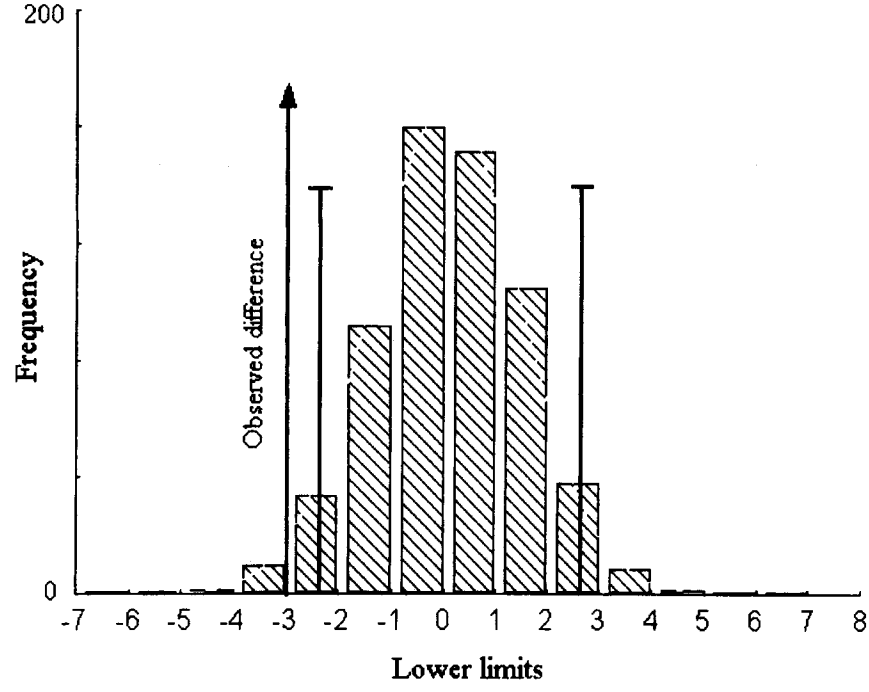

Figure 3. Histogram from 1,000 bootstrap resamplings on the difference between event-related potentials (Cz) to old and new low-imagery words for Subject A. Relative to this distribution, the observed difference (from the experiment) has a probability of less than .05.

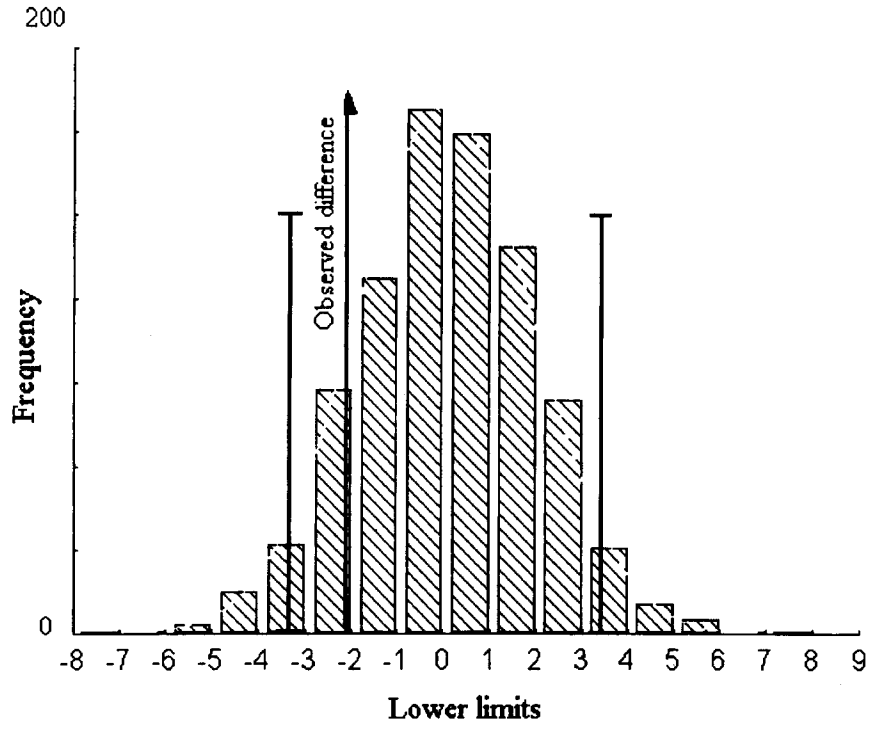

Figure 4. Histogram from 1,000 bootstrap resamplings on the difference between event-related potentials (Pz) to old and new low-imagery words. Relative to this distribution, the observed difference (from the experiment) has a probability of greater than .05.

(3.49 μV) than for HI words (about 1 μV; Table 3). There was no difference between ERPs to old and new words for subjects who completed the study phase with the interference task.

Bootstrap Analyses

The results showed that not all the subjects presented a reliable old/new effect; that is, the null hypothesis that the observed difference between mean amplitudes in the 400–800 msec latency range of ERPs to old and new items was due to chance cannot be rejected for all subjects (Table 4).

Figures 3 and 4 report the results of the bootstrap analysis applied to one reliable and one unreliable difference between ERPs to old and new words, showing the differences between the mean amplitudes of ERPs averaged on a random basis (bootstrap samples). In other words, the histograms represent the null hypothesis empirical distribution. As can be seen, in the first case, the frequency of occurrence of differences that are by chance equal or larger than the observed one is less than 50 out of 1,000 (which corresponds to a probability of .05). Hence in this case the observed difference between "real" ERPs can be said to be reliable. In the case of an unreliable effect, the probability associated with the observed difference is greater than .05 (Figure 4).

Thus, the conventional ANOVA showed a statistically significant old/new effect on ERPs in subjects from Group B, but the bootstrap analysis yielded a different picture (Table 4). In fact a reliable old/new effect, independent from the imagery value, was found only in 3 subjects out of 8 over Cz, and in 4 subjects out of 8 over Pz. No subject showed a reliable old/new effect over Fz.

In regard to ERPs to HI words, results showed that only 1 subject presented a reliable old/new effect over Fz and Cz, and 4 subjects presented a reliable effect over Pz. In regard to ERPs to LI words, the bootstrap analysis showed that no subject presented a reliable old/new effect over Fz, while 6 subjects presented the effect over Cz and Pz.

For the subjects in Group A, where a conventional ANOVA did not show any old/new effect, a reliable effect was indeed found over Pz (Table 4): In 3 subjects it was independent of imagery value, while a HI old/new effect was found in 1 subject and a LI old/new effect was found in 4 subjects.

Reliability of the old/new effect was not due to the specific value of the difference: In fact, the same difference can be reliable or not depending on the variability of the single-trial ERPs. For example, Figure 5 reports ERPs collected in a single subject from Cz and Pz leads. Although the differences between mean amplitudes of ERPs

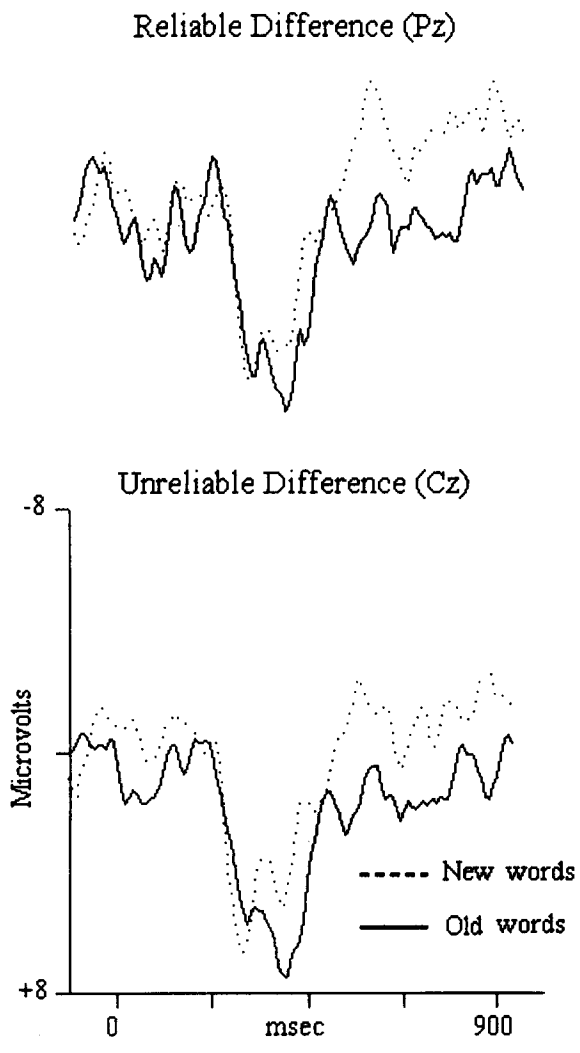

Figure 5. Event-related potentials to old and new words collected in a single subject from Cz and Pz leads. Bootstrap showed that only the difference over Pz is reliable. Stimulus onset at time 0. Negativity upward.

to old and new words were approximately the same over the two leads, only the difference over Pz was reliable according to the bootstrap analysis.

DISCUSSION

The aim of this study was to verify the usefulness of the bootstrap approach as a tool to obtain information about the reliability of experimental effects in cognitive neuroscience. The bootstrap technique was applied to ERP data from an experiment aimed at investigating the imagery-induced modulation of the old/new effect on ERPs recorded during a memory task.

The conventional ANOVA showed that subjects who completed an interfering orthographic task during the study phase did not present any ERP modulation due to imagery during the test phase, whereas subjects who completed the study phase without an interfering task showed an old/new effect on ERPs that was larger for LI than for HI words.

For all the subjects in the interference group, there was a difference between ERPs to old and new words, but bootstrap analyses showed that not all these differences were reliable. Depending on the electrode site, about half the subjects showed a difference that could be due just to chance. Reliable old/new effects were found mainly at the Pz lead and on ERPs to LI words. This uneven pattern confirms that bootstrap results depend on the experimental variables and are not due to chance. Furthermore, the reliability of the effect did not depend on its magnitude. In fact, the same difference can be reliable or not, depending on the EEG epochs entered in the averaging procedure.

Interestingly, whereas the ANOVA did not show any old/new effect modulation in subjects who completed an interfering task during the study phase, the bootstrap analysis showed that some of them indeed presented a reliable modulation. In summary, the bootstrap confirms the results from the ANOVA, but shows that the modulation of old/new effect is not very reliable, probably depending on the strategies subjects used to perform the task. It should also be noted that the bootstrap procedure is not aimed at giving indications about which factors affect the reliability of ERPs. It is only a method to assess whether a problem of reliability exists. For example, in this experiment, unreliable ERPs may derive from a latency jitter of the P300 component (which represents a violation of the assumption of signal invariance), from individual differences, from intrinsic intertrial variability of ERPs recorded during a recognition memory paradigm, or from something else. Factors determining the unreliable ERPs have to be identified by specifically designed studies.

These results confirm the usefulness of the bootstrap procedure and show some of its advantages. In fact, whereas in conventional ANOVAs all the ERPs are equally considered, regardless of their reliability, bootstrap performed for each single subject depends on all the available data; thus information about the reliability of the effects of experimental variables can be gained. In this way, results from conventional analyses can be interpreted more accurately.

The procedure proposed in this paper is not an alternative to conventional analyses, however, and any results from its application should be considered cautiously: They do not allow researchers to discard "unreliable subjects." The bootstrap can be used to check data for the presence of inconsistencies, but it cannot be used to establish which factor is involved.

REFERENCES

BARTOLINI, U., TAVAGLINI, C., & ZAMPOLLI, A. (1971). *Lessico di frequenza della lingua italiana contemporanea* [Usage norms for contemporary Italian]. Milan: IBM Italia.

BEISTEINER, R., HUTER, D., EDWARD, V., & KOCH, G. (1997). Brain potentials with old/new distinction of non-words and geometric figures. *Electroencephalography & Clinical Neurophysiology*, **99**, 517-526.

BENTIN, S. (1987). Event-related potentials, semantic process, and expectancy factors in word recognition. *Brain & Language*, **31**, 308-327.

BLAIR, R. C., & KARNISKI, W. (1993). An alternative method for significance testing of waveform difference potentials. *Psychophysiology*, **30**, 518-524.

DE WEERD, J. P. C. (1981). A posteriori time-varying filtering of averaged evoked potentials. *Biological Cybernetics*, **41**, 211-222.

DE WEERD, J. P. C., & MARTENS, W. L. J. (1978). Theory and practice of a posteriori "Wiener" filtering of averaged evoked potentials. *Biological Cybernetics*, **30**, 81-94.

DI NOCERA, F., & FERLAZZO, F. (1999). ERPs and cognition: Reliability and control [Abstract]. *Journal of Psychophysiology*, **13**, 207.

DI NOCERA, F., FERLAZZO, F., & GENTILOMO, A. (1996). Stabilité interindividuelle de l'effet de modulation de la composante P300 des potentiels évoqués cognitifs dans une condition de double tâche. *Psychologie Française*, **41**, 365-374.

DONALDSON, D. I., & RUGG, M. D. (1998). Recognition memory for new associations: Electrophysiological evidence for the role of recollection. *Neuropsychologia*, **36**, 377-395.

EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, **7**, 1-26.

EFRON, B., & TIBSHIRANI, R. J. (1993). *An introduction to bootstrap*. New York: Chapman & Hall.

FARWELL, L. A., & DONCHIN, E. (1991). The truth will out: Interrogative polygraphy ("lie detection") with event-related brain potentials. *Psychophysiology*, **28**, 531-547.

FERLAZZO, F., CONTE, S., & GENTILOMO, A. (1993a). Event-related potentials and recognition memory: The effect of word imagery value. *International Journal of Psychophysiology*, **15**, 115-122.

FERLAZZO, F., CONTE, S., & GENTILOMO, A. (1993b). Event-related potentials and recognition memory within the levels of processing framework. *NeuroReport*, **4**, 667-670.

FERLAZZO, F., & DI NOCERA, F. (1998). The serial position effect on ERPs recorded in a cued-recall task [Abstract]. *International Journal of Psychophysiology*, **30**, 226.

FERLAZZO, F., DI NOCERA, F., & DI SEGNI, S. (1998). Inside cognitive processes: Using ERPs to investigate the time course of the serial position effect in a cued recall paradigm. *General Psychology*, **1**, 155-168.

FRIEDMAN, D. (1990). Cognitive event-related potentials components during continuous recognition memory for pictures. *Psychophysiology*, **27**, 136-148.

FRIEDMAN, D., & SUTTON, S. (1987). Event-related potentials during continuous recognition memory. In R. Johnson, Jr., J. W. Rohrbaugh, & R. Parasuraman (Eds.), *Current trends in event related potentials research* (*Electroencephalography & Clinical Neurophysiology*, Suppl. 40, pp. 316-321). Amsterdam: Elsevier.

HUMPHREY, D. G., & KRAMER, A. F. (1994). Toward a psychophysiological assessment of dynamic changes in mental workload. *Human Factors*, **36**, 3-26.

JASPER, H. (1958). The 10-20 electrode system of the International Federation. *Electroencephalography & Clinical Neurophysiology*, **10**, 371-375.

JOHNSON, R., KREITER, K., RUSSO, B., & ZHU, J. (1998). A spatiotemporal analysis of recognition-related event-related brain potentials. *International Journal of Psychophysiology*, **29**, 83-104.

KARIS, D., FABIANI, M., & DONCHIN, E. (1984). "P300" and memory: Individual differences in the von Restorff effect. *Cognitive Psychology*, **16**, 177-216.

KARNISKI, W., BLAIR, R. C., & SNIDER, A. D. (1994). An exact statistical method for comparing topographic maps, with any number of subjects and electrodes. *Brain Topography*, **6**, 203-210.

NEVILLE, H., KUTAS, M., & SCHMIDT, A. L. (1982). Event-related potential studies of cerebral specialization during reading. *Brain & Language*, **16**, 300-315.

NOLDY-CULLUM, N. E., & STELMACK, R. M. (1987). Recognition memory for pictures and words: The effect of incidental and intentional learning on N400. In R. Johnson, Jr., J. W. Rohrbaugh, & R. Parasuraman (Eds.), *Current trends in event-related potentials research* (*Electroencephalography & Clinical Neurophysiology*, Suppl. 40, pp. 350-354). Amsterdam: Elsevier.

PAIVIO, A. (1965). Abstractness, imagery, and meaningfulness in paired-associate learning. *Journal of Verbal Learning & Verbal Behavior*, **4**, 32-38.

PRATT, H., EREZ, A., & GEVA, A. B. (1994). Lexicality and modality effects on evoked potentials in a memory scanning task. *Brain & Language*, **46**, 353-367.

RUGG, M. D. (1985). The effect of semantic priming and word repetition on event-related potentials. *Psychophysiology*, **22**, 642-647.

RUGG, M. D. (1995). ERP studies of memory. In M. D. Rugg & M. G. H. Coles (Eds.), *Electrophysiology of mind* (pp. 132-170). Oxford: Oxford University Press.

RUGG, M. D., FURDA, J., & LORIST, M. (1988). The effects of task on the modulation of event-related potentials by word repetition. *Psychophysiology*, **25**, 55-63.

RUGG, M. D., MARK, R. E., WALLA, P., SCHLOERSCHEIDT, A. M., BIRCH, C. S., & ALLAN, K. (1998). Dissociation of the neural correlates of implicit and explicit memory. *Nature*, **392**, 595-598.

RUGG, M. D., & NAGY, M. E. (1987). Lexical contribution to nonword-repetition effects: Evidence from event-related potentials. *Memory & Cognition*, **15**, 473-481.

RUGG, M. D., SCHLOERSCHEIDT, A. M., & MARK, R. E. (1998). An electrophysiological comparison of two indices of recollection. *Journal of Memory & Language*, **39**, 47-69.

SANQUIST, T. F., ROHRBAUGH, J. W., SYNDULKO, K., & LINDSLEY, D. B. (1980). Electrocortical signs of level of processing: Perceptual analysis and recognition memory. *Psychophysiology*, **17**, 568-576.

SCHLOERSCHEIDT, A. M., & RUGG, M. D. (1997). Recognition memory for words and pictures: An event-related potential study. *NeuroReport*, **8**, 3281-3285.

SMITH, M. E., & GUSTER, K. (1993). Decomposition of recognition memory event-related potentials yields target, repetition and retrieval effects. *Electroencephalography & Clinical Neurophysiology*, **86**, 335-343.

WASSERMAN, S., & BOCKENHOLT, U. (1989). Bootstrapping: An application to psychophysiology. *Psychophysiology*, **26**, 208-221.

WOODY, C. D. (1967). Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. *Medical & Biological Engineering*, **5**, 539-553.

(Manuscript received December 18, 1998;
revision accepted for publication October 3, 1999.)