Psychophysiology, 42 (2005), 356–359. Blackwell Publishing Inc. Printed in the USA.

DOI: 10.1111/j.1469-8986.2005.00283.x

Subjective and objective probability effects on P300 amplitude revisited

J. PETER ROSENFELD, JULIANNE R. BIROSCHAK, MELISSA J. KLESCHEN, AND KYLE M. SMITH
Department of Psychology, Northwestern University, Evanston, Illinois, USA

Abstract

Does objective probability affect P300 size independently and in addition to subjective probability? The latter was manipulated by the number of stimuli presented and classification task. Five groups saw target and frequent stimuli. Two saw these with $p = .2$ or .067, with two different button presses. Three groups saw two additional nontarget stimuli each with $p = .067$. One group pressed a different button for each stimulus. A second group pressed one button for the three oddballs, another for the frequent. A third *critical* group pressed one button for the target and another for other stimuli. In this group, P300 was larger for targets versus nontargets, and larger for nontargets versus frequents. Although nontargets were classified with frequents, their actual low probability distinguished them from frequents, and their subjective probability distinguished them from targets. Therefore, actual and subjective probability effects were independently found.

Descriptors: Event-related potential, P300 amplitude, Subjective probability, Stimulus classification

In one of many reports on the influence of probability on P300 amplitude, Duncan-Johnson and Donchin (1977) wrote, "It is reasonable to assume that experimental manipulations of the a priori probability of stimuli can affect the amplitude of the P300 only to the extent that they have an effect on the subjective probability of the event" (p. 456). This view was supported by a study in which actual probability of counted tones was experimentally manipulated, and P300 averages were computed as a function of the preceding sequence of counted stimuli. The typical inverse relationship between a priori probability and P300 amplitude was obtained, however, with the slope of the declining function greatly reduced when the stimulus was the first in a series. Moreover, *frequent* counted stimuli (probability of .7 to .9) that were the first in a series elicited sizeable P300s. Humans "seem to find it exceedingly difficult to accept the fact that in a random series of events the probability of any outcome on trial *n* is independent of the outcome of the previous trials," (p. 456) that is, objective or actual probability was played down as a key influence on P300 amplitude, and subjective probability was emphasized. The same ultimate conclusion was reached by Karis, Chesney, and Donchin (1983) via manipulation of payoff structure. As summarized by Donchin, Karis, Bashore, Coles, and Gratton (1986), "subjective rather then objective probability . . . is critical for the elicitation of P300" (p. 249).

Johnson and Donchin (1980) provided dramatic evidence that P300 amplitude is inversely related to the *subjective* and not the *objective* probability of the target. They described an experiment having three tonal stimuli, each with the objective probability of .33. The instructions were to count occurrences of one stimulus, but not the other two. This forced two stimuli into a single, frequent, nontarget category, and the counted stimulus became a rare or oddball stimulus, which elicited a larger P300 response. In a two-stimulus condition (Johnson & Donchin, 1980), in which one counted target had a probability of .33 and an uncounted frequent had a probability of .67, the difference in P300 amplitude was not larger than that seen in the three-stimulus condition. Johnson and Donchin concluded: "Even though all *stimuli* in this (three stimulus) condition occurred equally probably, the ERPs elicited by the non-target stimuli suggest that the subject treated each uncounted stimulus as if its prior probability were .67—that is *twice* its actual value" (p. 170).

Although there is indeed strong evidence for the influence of subjective probability in these relatively high probability, simple stimulus situations (e.g., that include only two or three stimuli), the conjoint and independent influence of objective as well as subjective probability may be detected in a more complex design. In the current study, intended as an extension of Johnson and Donchin, (1980), we used four stimulus types and lower oddball probabilities, and we also investigated the influence of task demand (Kramer, Sirevaag, & Braune, 1987). Thus we explored the independent effects of three factors, subjective probability, objective probability, and task demand, in one experiment.

In this design, a critical condition, designed to be an extended version of the Donchin–Johnson (1980) three-stimulus paradigm, is a task with four visual stimuli, each a distinct letter triad. Three stimuli are rare with a probability of .067 (25 of 375 total trials); one is frequent (probability of .8, 300 trials). The subject presses Button 1 to one rare target and Button 2 to the other two

Address reprint requests to: J. Peter Rosenfeld, Department of Psychology, Northwestern University, Evanston, IL 60208, USA. E-mail: jp-rosenfeld@northwestern.edu.

rare nontargets and to the frequent stimulus. It is expected that the target will evoke the largest response, because it is the rarest stimulus category, and because it requires a unique response, that is, it has the lowest objective *and* subjective probabilities. However, it is also expected that the two rare nontargets will evoke a smaller P300 than that evoked by the target, but larger than that evoked by the frequent. This relation is expected because the two rare nontargets have smaller *objective* probabilities, relative to the frequent, even though they are categorized together with the frequent stimuli.

In four other comparison groups we varied the number of stimuli presented (two or four) and the number of response buttons (two or four), which together define different subjective and objective probabilities, as well as different task demands. The goal is to determine whether task demand interacts with categorization and objective probability.

Methods

Participants

Participants were 56 members of an Introductory Psychology course at Northwestern University who had normal or corrected-to-normal vision. All signed an IRB-approved consent form.

Design

See Table 1 for a design summary. In each of five groups, participants saw stimuli presented, one every 3 s, on a video display. All groups had a four-button box mounted as an extension of the right arm rest. The first two groups saw two stimuli in a classic oddball task with differing target probabilities, .2 (ODD/.2; $n = 10$), or .067 (ODD/.067; $n = 9$). The *target* stimulus requiring a Button 1 press was the letter triad, DDD. The *frequent* triad (Button 2 press) was XXX.

In addition to these two stimuli, participants in each of the remaining three groups saw two additional rare, *nontargets*, MMM and RRR, both with a probability of .067 (which is the same as DDD). These *four stimulus* (*4S*) groups differed in terms of which buttons were pressed to the triads. Group 4S1112 ($n = 12$) pressed Button 1 to the target and to both nontargets and Button 2 to frequents. Categorization of targets and nontargets into one response category was expected to give these stimuli a conjoint probability of .2, each evoking a P300 resembling that of the target in group ODD/.2. Group 4S1222 ($n = 13$) pressed Button 1 to targets and Button 2 to the other stimuli. Group 4S1234 ($n = 12$) pressed Buttons 1, 2, 3, and 4 to the target, nontarget1, nontarget2, and frequent, respectively, and this task was expected to be the most demanding with its unique four-response requirement. Also, although each stimulus had its own response button, the actual probability of .8 for the frequent would reduce its P300 despite categorization.

Participants sat in a recliner; the display face was 1 m from their eyes.

Data Acquisition and Analysis

EEG was recorded with silver electrodes at sites Fz, Cz, and Pz referenced to linked mastoids. EOG was recorded with silver electrodes above and below the right eye. The artifact rejection criterion was 80 μV. The EEG electrodes were referentially recorded but the EOG electrodes were differentially amplified. The forehead was grounded. Signals were passed through Grass P511K amplifiers with a 30-Hz low-pass filter setting, and with high-pass filters set (3 db) at 0.1 Hz. Amplifier output was passed

Table 1. *Designated Button Presses, Objective Probabilities, Numbers of Trials, RTs, and Error Rates for All Stimuli and Groups*

Group	Target	Nontarget1	Nontarget2	Frequent
ODD/.2				
Button press	1	—	—	2
Number of trials	75	—	—	300
Probability	.2	—	—	.8
Error rate	.006	—	—	.000
RT	441			375
ODD/.067				
Button press	1	—	—	2
Number of trials	25	—	—	350
Probability	.067	—	—	.933
Error rate	.073	—	—	.000
RT	519			378
4S1234				
Button press	1	2	3	4
Number of trials	25	25	25	300
Probability	.067	.067	.067	.8
Error rate	.012	.056	.045	.000
RT	672	743	754	471
4S1222				
Button press	1	2	2	2
Number of trials	25	25	25	300
Probability	.067	.067	.067	.8
Error rate	.028	.000	.000	.000
RT	606	556	575	453
4S1112				
Button press	1	1	1	2
Number of trials	25	25	25	300
Probability	.067	.067	.067	.8
Error rate	.000	.017	.018	.000
RT	527	542	521	433

to a 12-bit A/D converter sampling at 125 Hz. For all analyses and displays, single sweeps and averages were digitally filtered off-line to remove higher frequencies; 3 db point = 4.23 Hz. P300 was measured using the base to peak method (BASE–PEAK): The algorithm searched within a window from 400 to 900 ms for the maximally positive segment average of 104 ms. The prestimulus 104-ms average was also obtained and subtracted from the maximum positivity to define the BASE–PEAK measure.

Results

Behavioral Performance

Error rates, listed in Table 1, were low overall. Dropping the frequent condition (having no errors), a three stimulus × three group ANOVA on error rates showed a main effect of group approaching significance $F(1,34) = 3.98$, $p = .054$, apparently carried by the higher error rate across stimuli in group 4S1234, compared to groups 4S1222 and 4S1112, respectively, consistent with the hypothesis that the four-button task in group 4S1234 was the most demanding. The error rate for targets was also somewhat higher for group ODD/.067, suggesting difficulty switching responses so rarely. Error rates in groups ODD/.067 and 4S1234 did not differ, $t(8.5) = 1.22$, $p = 0.26$.

Table 1 lists reaction times (RTs). As expected, RT appears elevated to the low probability stimuli in group 4S1234. Groups 4S1222 and 4S1112 do not appear to differ. Among these three groups viewing four stimuli, two orthogonal ANOVAs were conducted using the target, nontarget1, and nontarget2 stimuli: Group 4S1222 did not differ from group 4S1112, as expected, and these combined groups differed greatly from group 4S1234. The group effect was large, $F(1,34) = 32.9$, $p < .001$. Consistent with the error data, these results support the expectedly greater task demand in group 4S1234. The easiest task appears to have been for group ODD/.2; a post hoc comparison showed faster RT when compared to groups 4S1222 and 4S1112 yields $t(24.9) = 4.64$, $p < 0.001$.

P300 (Pz) Amplitude

Figure 1A illustrates grand averages by group and stimulus. Planned comparisons were employed to test specific predictions.

The classic effect of target oddball probability was obtained with target P300 amplitude (or with target minus frequent difference) in group ODD/.067 larger than that in the other groups, $t(54) = 2.99$, $p < .005$, which did not differ from each other. In this group, there is a simple, two-stimulus task with the rarest oddball.

In groups 4S1234 and 4S1112, there is no difference between the P300s to target and nontarget stimuli, whereas in group 4S1222, the response to nontarget1 appears to be midway between those of the target and the frequent. These impressions are also illustrated in Figure 1B.

Group 4S1222 was expected to show both subjective and objective probability effects. Rare nontargets were expected to show larger P300s than frequent stimuli, but less than for rare targets. As expected, the two rare nontargets were not different, as these had the same objective and subjective probabilities, .067. As hypothesized, however, P300 amplitude was larger for rare targets compared to rare nontargets, $t(11) = 3.18$, $p < .01$, and P300s for rare nontargets were larger than for frequent stimuli, $t(11) = 3.97$, $p < .003$.

Groups 4S1234 and 4S1112 were expected to have nontarget P300s greater than those of group 4S1222 because only in group

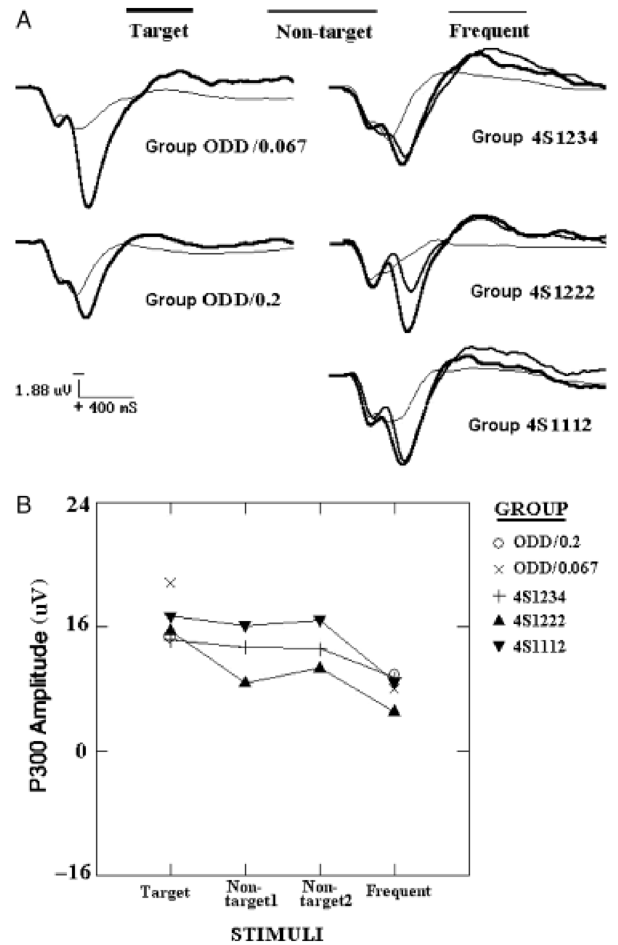

Figure 1. A: Grand averaged Pz P300s to targets, nontarget1, and frequents; all groups. B: Line graphs of computed Pz P300 amplitude averages within each group for the four stimuli.

4S1222 did categorization by button press force responses to nontargets into the same subjective probability category as frequents. The mean of nontargets was greater in groups 4S1234 and 4S1112 combined than in 4S1222, $t(35) = 2.66$, $p < .02$. We also expected that 4S1234 would have smaller P300s to nontargets, relative to 4S1112, because 4S1234 had the greater demand. Although Figure 1B suggests that nontargets were larger in group 4S1112 than in 4S1234, a t test comparing nontarget means between these groups was not significant, $t(23) = 1.41$, $p = .17$. Demand effects were also illustrated by comparing (post hoc) target amplitude between groups ODD/.067 and 4S1222: Both had target probabilities of .067 and only two buttons, but 4S1222 saw more stimuli. Thus, for ODD/.067 compared to group 4S1222, $t(19) = 3.061$, $p < 0.007$, with the latter group having the smaller P300. Post hoc comparison of target RTs between these groups is consistent: $t(19) = 2.32$, $p < 0.03$ with 4S1222 having the longer RT.

To confirm the objective probability effect, although each stimulus had its own response category (subjective probability), we compared the mean of responses to combined targets and nontargets in group 4S1234 with those to frequents: Targets and nontarget mean was larger, $t = 3.302$, $df = 11$, $p < .008$. Finally, consistent with the prediction that target P300s would be the same in groups ODD/.2 and 4S1112, the contrast of those groups yielded a nonsignificant $p > .3$.

Discussion

The literature of 1970–1990 regarding probability and P300 amplitude emphasized subjective, not actual probability. We hypothesized that if oddball stimuli were very rare, categorization manipulations would fail to eliminate actual probability effects. As shown in group 4S1222, the order of amplitude effects was that P300 was larger for targets compared to nontargets, and these latter P300s were larger than those for frequents: Targets, with actual probability of .067 and requiring a unique response, produced the largest P300; frequents, with probability of .8, produced the smallest. Nontargets, each with probability of .067 (equal to target probability), were categorized with frequents, and so produced amplitudes smaller than the target P300, but their actual low probability produced an oddball effect distinguishing them from frequents. Although the actual probabilities of nontargets were the same in all four-stimulus groups, in 4S1234 and 4S1112, nontargets were not categorized with frequents, so that their actual probability effects were not pulled down by categorization. Indeed in 4S1234, when categorization alone would have produced four categories, there was a distinct actual probability effect of combined targets and nontargets versus frequents.

Mecklinger and Ullsperger (1993) reported a related study in which the stimuli were five words for the numbers one through five. The key difference was that all their stimuli were equally probable at 0.2. They found that unique response categorization elevated P300 amplitude to any designated target, but did not find objective probability effects in interaction. In the current study, manipulated subjective and objective probability did interact.

In Duncan–Johnson and Johnson (1977), although the decreasing function of P300 amplitude and actual probability was modulated by sequence, the slope remained negative across differing sequences. The present oddball effects of nontargets in group 4S1222 ultimately must be subjective effects because human subjects must subjectively perceive the actual probabilities. This suggests a restated conceptualization of probability effects: It *is* subjective versus simple objective probability that influences P300. Subjective probability, however, may be influenced by many factors, including (but not necessarily limited to) stimulus categorization (Johnson & Donchin, 1980), sequence (Duncan-Johnson & Donchin, 1977), payoff matrix (Karis et al., 1983), interstimulus interval (not yet noted here, but see Polich, 1990, and Polich & Bondurant, 1997), and perhaps most basically, objective probability.

A feature of the studies reported and cited by Johnson and Donchin (1980) was the analysis of sequential effects in the ERP data, that is, a target-evoked P300 was larger if preceded by a run of nontargets than if preceded by a target run. Given the constant target probability over the entire set of trials, these sequential data clearly illustrated how local sequences influence subjective probability and P300 amplitude. Our design focused on very low probability effects and thus demanded very few target and rare nontarget events, which precluded a sequential analysis, because our relatively large number and associated high density of frequent trials greatly reduced the possibility of target and rare nontarget trial runs. In addition, Johnson and Donchin used shorter duration (acoustic) stimuli and intertrial intervals, resulting in many more trials overall (2225 vs. 375 here). This also precluded our doing a meaningful sequential analysis in the current design. Because our interest was in *objective* probability for events with low probability, we were able to demonstrate these effects without studying sequence data, which, after all, have been directed mainly at illustrating *subjective* effects. Our study clearly showed the independent effects of subjective and objective probability within one experiment and, indeed, within one group (4S1222).

REFERENCES

Donchin, E., Karis, D., Bashore, T. R., Coles, M. G. H., & Gratton, G. (1986). Cognitive psychophysiology and human information processing. In M. G. H. Coles, E. Donchin, & S. W. Porges (Eds.), *Psychophysiology: Systems, processes, and applications* (pp. 244–266). New York: Plenum Press.

Duncan-Johnson, C. C., & Donchin, E. (1977). On quantifying surprise: The variation of event related potentials with subjective probability. *Psychophysiology*, *14*, 456–467.

Johnson, R. Jr., & Donchin, E. (1980). P300 and stimulus categorization: Two plus one is not so different from one plus one. *Psychophysiology*, *17*, 167–178.

Karis, D., Chesney, G. L., & Donchin, E. (1983). "... 'Twas ten to one; And yet we ventured ..." P300 and decision making. *Psychophysiology*, *20*, 260–268.

Kramer, A. F., Sirevaag, E. J., & Braune, R. (1987). A psychophysiological assessment of operator workload during simulated flight missions. *Human Factors*, *29*, 145–160.

Mecklinger, A., & Ullsperger, P. (1993). P3 varies with stimulus categorization rather than probability. *Electroencephalography and Clinical Neurophysiology*, *86*, 395–407.

Polich, J. (1990). P300. Probability, and interstimulus interval. *Psychophysiology*, *27*, 396–403.

Polich, J., & Bondurant, T. (1997). P300 sequence effects, probability, and interstimulus interval. *Physiology and Behavior*, *61*, 843–849.

(Received August 5, 2004; Accepted January 20, 2005)