Absence of the gambler's fallacy in psychophysical settings*

HERBERT A. COLLE†

University of Chicago, Chicago, Illinois 60637

RICHARD M. ROSE

University of Washington, Seattle, Washington 99195

and

H. A. TAYLOR

Rutgers University, New Brunswick, New Jersey 08903

The gambler's fallacy, or the negative recency effect, occurs on the early trials of binary prediction experiments and has been attributed to the problem-solving activities of Ss in these situations. In two different experiments, no evidence of a negative recency effect was found for Ss' judgments in two-choice recognition experiments, although the typical negative recency effect was found when the same sequences of outcomes were used in binary prediction tasks. The data suggest that the problem-solving attitude is responsible for the negative recency effect. The absence of negative recency effects also suggests that simple path-independent models may provide adequate descriptions of response bias changes in judgmental tasks.

If people are asked to predict whether a sequence of like events will continue or not, they are more likely to predict that the current event will be repeated after seeing a short run of these events than after seeing longer runs (Brown & Overall, 1959; Bush & Morlock, 1959; Derks, 1962; Edwards, 1961; Estes, 1957; Friedman, Burke, Cole, Keller, Millward, & Estes, 1964; Lindman & Edwards, 1961; Nicks, 1959). This negative recency effect has been called the gambler's fallacy. It was theoretically important because it contradicted a parameter-free prediction of simple path-independent linear operator and stimulus sampling models (Brown & Overall, 1959; Bush & Morlock, 1959). These models predict that after the occurrence of an outcome event, the probability of predicting another occurrence should increase or at least should not decrease. That is, the models predict positive recency. The existence of the negative recency effect is one of the reasons that most current theories for performance in binary prediction experiments either assume that Ss are behaving like problem solvers or, at least, that models more complicated than the path independent ones are needed (Jones, 1971).

Several models have been proposed to describe the sequences of judgments made by Ss in psychophysical experiments (see Bush, Luce, & Rose, 1964; Friedman, Carterette, Nakatani, & Ahumada, 1968; Schoeffler, 1965). These models have incorporated into them

assumptions similar to the assumptions of the simple path-independent models that were suggested for binary prediction experiments. In the models discussed by Friedman et al (1968), on a proportion of the trials, the S's responses are determined by the differences between the stimuli. However, on the remainder of the trials, the Ss are forced to guess which of the two responses is correct. These models assume that this guessing probability is determined by the sequence of feedback events that are provided to inform the S which responses were correct.

In the other models, a response bias parameter is adjusted by a learning operator after each trial's outcome is presented. In all cases, a positive recency effect is predicted. That is, the predicted probability of Stimulus j after k presentations of Stimulus j is a nondecreasing function of k .

The occurrence of the negative recency effect in binary prediction experiments does not imply that the same effect must be found for S's judgments in psychophysical settings, although it is a part of laboratory lore that it does occur, and some investigators warn their Ss about it. For example, to generate a negative recency effect, it may be necessary for Ss to adopt a problem-solving attitude towards a guessing task. Ss in psychophysical experiments probably are concentrating on judging stimuli and are not trying to solve a problem that depends on knowing the past history of outcomes. The past history of outcomes may have an influence on future responding, but it may be explained by a simple path-independent biasing mechanism.

There are at least two reasons why a serious test of the recency prediction is difficult in psychophysical experiments. One problem is the magnitude of the predicted effect. In the Friedman et al (1968) models, if the proportion of correct responses is 0.75, then, as a

*This research was supported in part by Grant MH 17761-01 from the National Institute of Mental Health, by funds from the Department of Psychology of the University of Washington, and by the Social Science Divisional Research Fund, University of Chicago. We wish to thank W. K. Estes, M. H. Smith, Jr., and J. Egan for the loan of equipment.

†Requests for reprints should be sent to Herbert A. Colle, Department of Psychology, Wright State University, Dayton, Ohio 45431.

rough approximation, the Ss are detecting stimulus differences on about 50% of the trials (assuming both stimuli are presented equally often). Since stimulus differences are detected independently of the sequences of feedback events, these stimulus differences would cause about a 50% reduction of any expected effect. A similar argument can be made for the other models.

A second problem deals with a possible interpretation of the negative recency effect rather than an implication of model assumptions, and is somewhat harder to explain. Most recency curves from binary prediction experiments suggest that the probability of predicting a run continuation increases somewhat after the first two or three consecutive events before decreasing after longer run lengths. In prediction experiments, Ss must guess on each trial, so it is not clear whether the negative recency effect is caused by the S's experiencing k consecutive occurrences of one outcome event and then guessing, or whether it is necessary that the S guess at k consecutive occurrences of one outcome event and then guess again. In psychophysical experiments, this distinction is important because the Ss do not necessarily guess on each trial. The models mentioned above do make a similar distinction. For some, the guessing probability changes after every outcome event; for the others, the guessing probability changes only after trials on which the S was forced to guess. The distinction is important, because under normal testing conditions the Ss would only very infrequently make more than two or three guesses in a row. Thus, the Ss would seldom experience the situation of having guessed at 4, 5, or 6 consecutive outcomes of one event. Hence, even if the recency curve found in binary prediction experiments were appropriate for psychophysical guessing, almost all of the guessing trials would occur under conditions that hold for the early part of the recency curve, which is increasing, and the decreasing part of the recency curve would be unlikely to be detected.

To minimize these problems, experiments were conducted in which the Ss ostensibly were to identify the presented stimulus as one of two possible stimuli, but actually no stimulus differences were present. Since the stimuli were constant from trial to trial, the experiments were formally identical to binary prediction experiments. Because a stimulus to be judged did occur on each trial, the Ss may have experienced differences which they reported, but these nonguessing trials should have occurred only infrequently.

It is important, if the results of the present experiments are to be generalized to more typical psychophysical experiments, that the Ss approach the task as they would have if stimulus differences had been presented. To foster the proper attitude, the Ss were given some trials with detectable stimulus differences at the start of the experiment and were told to expect the differences to become smaller over trials. Comparable data from binary prediction experiments also were collected for comparison.

EXPERIMENT I

Method

Subjects

Twenty students at Stanford University served as Ss in the visual recognition task.

Apparatus

The experimental apparatus used was an automated tachistoscope at the Institute for Mathematical Studies in the Social Sciences, Stanford University (see Taylor, 1966, for detailed description). The display terminus of the apparatus was located in an air-conditioned room, separate from the control equipment. Two observation stations were located 5½ ft apart, at opposite ends of a box containing projection equipment. At each station, a ground-glass screen was mounted in a recess in the box, centered 10 in. behind a plastic eyepiece. Stimuli were projected onto the screen through an aperture providing an illuminated circle of 2-in. diam; during the experiment, this was the only source of light in the room. Stimuli were displayed by a random-access projector located inside the box. Responses were indicated by pushing one of two buttons on a panel at lap height. The two outcomes were indicated by small lights mounted behind the screen; these were located above the illuminated circle, 7 in. apart on opposite sides of the screen.

Three types of display were used. Their basic form was that of a letter "X," composed of two line segments intersecting at their midpoints. Display D_0 was formed by aligning the sides of the "X" at exact 45-deg angles to the horizontal, so that the angles of intersection of the line segments were all 90 deg. Displays D_1 and D_2 were formed by rotating both segments 1 deg in opposite directions about the point of intersection, so that the resultant "X" included angles of 88 and 92 deg. Display D_1 denotes the figure with the lesser angles at the top and bottom, and D_2 that with the greater angles at the top and bottom.

The displays appeared as black figures against a white background. In all, 48 slides were used, 16 copies of each of the three displays. The "X" was centered on each slide. When projected on the screen, the line segments appeared 5/8 in. long and 1/16 in. thick. At a viewing distance of 11 in., this figure would subtend a visual angle of approximately 2 deg 18 min.

Procedure

Each S was run for a single 1-h session. Ss were run in pairs. At the start of the session, each S adjusted the height of his chair so that he could view the screen through the eyepiece without strain. The instructions were then read; they emphasized the psychophysical nature of the experiment and encouraged accuracy. Ss were asked to indicate on each trial whether they thought a display of Type D_1 or Type D_2 had been presented. The response buttons were labeled with exaggerated forms of Displays D_1 and D_2 to show how decisions were to be indicated. Ss were told that if the outcome light following a response was on the same side of the screen as the button pushed, the response was correct; otherwise, the response was incorrect. Correct responses were to be rewarded with a payoff of 1 cent, and incorrect responses led to the loss of 1 cent.

Each session consisted of the trials of the following format: The trial began when the projector advanced to the slide to be presented. Selection time averaged 1.2 sec and provided no consistent auditory cue to the type of slide selected. Two seconds from the start of the trial, a click was delivered over the intercom; ½ sec later, the slide was presented for 25 msec. No limit was placed on the time allowed for response. After both Ss had responded, the outcome light was lit for 3.5 sec.

Two types of trial were used, practice and bias trials. On

practice trials, a D_1 or D_2 slide was presented, and the outcome was consistent with the stimulus. On bias trials, a D_0 slide was always presented, and the outcome was E controlled. The presentation schedule started with two blocks of 50 practice trials; each block was an irregular ordering of 25 D_1 and 25 D_2 trials. These were followed by a block of 250 bias trials; Outcomes O_1 and O_2 occurred on 75% and 25% of these trials, respectively, in an irregular order identical to the one used by Bush and Morlock (1959).

There was no formal rest break, but Ss were permitted to take short breaks by delaying responses if they so desired. After the session, Ss were asked whether they had responded throughout the experiment on the basis of the stimulus presentations.

Results

We consider only runs of the more frequent event. Let F_k denote a prediction by the S that the run will continue after the first k events of the run. The data from this experiment are presented in the right panel of Fig. 1.¹ The proportion of F_k events is plotted as a function of k , the length of the ongoing run of the more frequent event. Run length zero denotes the occurrence of the less frequent event without regard to run length. The horizontal bars represent 95% confidence intervals about the data points. The confidence intervals are wider for the longer run lengths, because the proportions are based on a smaller number of observations. From left to right, the proportions are based on 63, 50, 37, 33, 22, 14, 11, and 9 predictions by each S.

For comparison, the left panel presents the same data from the experiment conducted by Bush and Morlock (1959). Bush and Morlock's experiment was a standard binary prediction experiment. The visual recognition experiment described above used a sequence of outcome events identical on each trial to the sequence used by Bush and Morlock.

In the data from the visual recognition experiment (right panel), there is no evidence of a significant decrease in any part of the curve, i.e., no evidence of a negative recency effect. In contrast, the Bush and Morlock data (left panel) exhibit the typical negative recency effect. The proportion of F_k events clearly decreases for values of k greater than two.

EXPERIMENT II

Method

Subjects

For the auditory recognition task, 20 Ss were obtained by advertising on the University of Washington campus. The Ss were paid \$2 for the 40-min session. For the binary choice task, 100 undergraduates at the University of Washington participated to fulfill part of a course requirement.

Apparatus for the Auditory Recognition Task

The stimuli for the auditory task were generated by switching the output from an oscillator into one of three attenuators. One of the attenuators controlled the intensity of the standard tone; the other two attenuators were set at 1.0 dB above and below

Fig. 1. Proportion of predictions of the more recent event, F_k , as a function of the run length of that event.

the attenuation of the standard intensity tone. A relay with mercury-wetted contacts was used to gate the signal. Switching transients were reduced by passing the signal through a low-pass section of a Spencer-Kennedy electronic filter, which provided 18 dB per octave attenuation above the cutoff frequency of 1,000 Hz. The signal was amplified and sent to three Permaflux PDR-8 earphones. Each earphone was located in the left earmuff of a pair of Wilson Sound Barrier earmuffs. The standard tone had a frequency of 1,000 Hz and an intensity of 85 dB SPL, which was comfortable and clearly audible.

The testing was conducted in three single-walled Industrial Acoustics sound-attenuating booths (Model 400). The display panel in each booth had a white warning light, centered at the top, two green circles to mark the observation intervals, four response levers, and four small red lights just above each response lever. The Ss used only the middle two levers; the outermost levers were nonfunctional. The left lever was marked "softer," and the right lever was marked "louder." The light just above each response lever was lit when the corresponding lever was pressed during the response interval. The two end lights were used to present feedback to the Ss.

Procedure for the Auditory Recognition Task

The Ss were tested in groups of three or fewer. Ostensibly, their task was to identify the tone presented on each trial as one that was either more or less intense than the standard intensity tone. The standard tone was presented three times before the experiment started. Each trial consisted of a warning interval, an observation interval, a response interval, a feedback interval, and an intertrial interval. The duration of each of these intervals was 0.5 sec, except for the response interval, which had a 2.0-sec duration.

The Ss were encouraged several times in the instructions to pay close attention to the signals for the full 40 min of the experiment, and were offered a \$1 incentive contingent on their obtaining 60% or more of their responses correct. The Ss were instructed to guess if they were not certain of their choices, and were told that the feedback was there to help them make their decisions. In addition, the Ss were warned that the tones were difficult to discriminate and that they would become more difficult as the experiment progressed.

For the first 50 trials, the Ss received an irregular sequence of 25 trials in which the signal tone was 1.0 dB less intense than the standard and 25 trials in which the signal tone was 1.0 dB more intense than the standard. The presentation of the feedback lights was consistent with these differences. For the next 503 trials, the standard tone itself was presented on each trial and the presentation of the feedback lights was determined by one of the random sequences.

Nicks's (1959) procedure was used to generate the random

Fig. 2. Proportion of predictions of a run continuation, Y_k , as a function of run length for the first 250 trials.

sequences. The probability of each feedback event's occurring was set at 0.5, and the expected number of runs of each length for samples of 500 trials was calculated. A population of run lengths was constructed in which the number of runs of each length approximated the expected number. The two distributions of run lengths (one for each event) were divided in half, so that blocks of 250 trials would have approximately the same run structure. Finally, the random sequences were generated by sampling without replacement from these distributions until they were exhausted. Two different random sequences were generated in this way. Both sequences were started equally often with each feedback event, and an equal number of Ss were tested with each of the two sequences.

Apparatus for the Binary Choice Task

The apparatus consisted of a long table divided into five booths by curtains. Each booth was equipped with a pair of keys which the S used to make his predictions. All five booths faced a single 19-in.-wide $\times$ 14-in.-high display panel, located about 10 ft across the room from the Ss. A yellow warning light was located 9.5 in. from each end of the panel. Two red event lights were located slightly below the warning lights and 5.5 in. from the center of the panel.

The sequence of trial events was programmed to run automatically. The programming equipment was located in the same room as the Ss, but it was activated only after all the Ss had responded. This procedure eliminated possible extraneous cues from the equipment.

Procedures for the Binary Choice Task

The Ss were tested in groups of five or fewer. They were instructed that, when the yellow warning came on, they were to predict which of the two event lights would come on. The left key was to be pressed to predict the occurrence of the left light, and the right key was to be pressed to predict the occurrence of the right light. The trials were group paced; the appropriate event light came on 1.4 sec after all the Ss had responded. It stayed on for 0.5 sec and was followed by an 0.5-sec interval before the warning light came on again.

Besides specifying the details of the procedure, the instructions contained the information that the experiment was a learning experiment and that the Ss were to predict the events which would occur. They were told that the sessions were conducted with groups rather than individuals for convenience only and that the other Ss' presence was irrelevant for the task. The Ss were told nothing about the sequences of events. The sequences that were used were the same as those used for the auditory task. All Ss were asked to try hard because the

experiment was a test of the limitations of the human learning process. An equal number of Ss were tested with each of the two sequences, and each sequence started an equal number of times with each event.

Results

The data for each of the two events were combined because the results of each event were similar and because the distribution of runs was the same for each event. Let Y_k denote a prediction by the S that the run will continue after the first k events of the run. Figure 2 presents the proportion of Y_k events as a function of k for the first 250 trials. As before, the left panel presents the recency curve from the binary choice task and the right panel presents the recency curve from the recognition task. The horizontal bars represent 95% confidence intervals about the points. From left to right for both tasks, the proportions are based on 125, 63, 32, 16, 8, and 4 observations per S. Notice that the recency curve from the binary choice task generally decreases with increasing run length, i.e., there is an increasing tendency for Ss to predict that a run will end as the length of the run increases. This result is the usual negative recency effect. The curve from the recognition task in the right panel has a different form. It increases up to a run length of three. The confidence intervals for the last three points are so wide, it is impossible to tell if the curve then levels off or decreases.

Figure 3 presents the recency curves from the second 250 trials. The curve from the binary prediction task in the left panel now has a clear initial rise before it drops. Although the curve drops below its maximum, it never drops below its initial level as the curve from the first 250 trials does. Edwards's (1961) results show a similar reduction in negative recency effects over trials.

Once again, looking at the right panel, the recognition data show no clear evidence of a negative recency effect. In addition, the probability of predicting a run continuation increased following the longer run lengths.

DISCUSSION

The existence of the negative recency effect was one of the original reasons for abandoning the use of simple path-independent models to describe performance in binary choice experiments. Violations of the positive recency prediction were found, although the models seemed to be performing adequately when tested using other statistics (Friedman et al, 1964). In particular, sequential statistics based on only the past one or two trials usually were consistent with model predictions. In testing the corresponding models in psychophysical settings, again sequential statistics based on only the past one or two trials were not grossly in error (Atkinson, 1963; Atkinson & Kinchla, 1965). However, the recency curves were not reported for these situations. Since sequential statistics based on one or two previous trials are heavily weighted with short runs, they could be

consistent with recency curves which increase after several successive events and then decrease. In the present experiments, we found that for a visual and for an auditory recognition experiment, the recency curves were consistent with the positive recency prediction. Therefore, the simple path-independent models have passed a severe test in psychophysical settings which they did not pass in binary choice settings.

It is interesting to consider the nature of the differences between the way Ss operate when making psychophysical judgments and when making predictions in the absence of any stimuli to judge. The above evidence suggests that in psychophysical tasks with information feedback, Ss' performance satisfies the path-independence assumption, which is consistent with Ss concentrating on comparing the perceptual information with the outcome information presented on that trial and ignoring the pattern of results of the immediately preceding trials. The changes in response probabilities that follow the occurrence of different sequences of outcome events are consistent with the existence of a simple biasing mechanism which is isolated by focusing the S's attention on the stimulus. Recognition tasks may be particularly useful for fostering this attitude, since a clearly perceived stimulus to be judged is presented on each trial. In "absolute threshold" experiments in which minimally detectable stimuli are used, Ss may be less likely to adopt the judgmental attitude if there is a large number of trials on which nothing or very little is perceived when the stimulus is presented. The lack of a judgeable perception may lead the Ss to adopt a guessing strategy or problem solving set towards the task.

Ward and Lockhead (1970, 1971) have run a series of studies of absolute judgment of easily perceived stimuli. Their studies are similar to those reported above in that they have compared a guessing or probability learning situation (1971, Study 3) with a difficult psychophysical situation. They have examined sequential effects in absolute judgment experiments in which Ss attempted to identify which of 10 stimuli had been presented, and then compared these to the sequential effects found when Ss predicted which one of 10 digits would be presented next. They report two types of sequential effects to be present in those conditions in which they gave their Ss feedback. The larger of these effects is a tendency to respond with responses corresponding to the stimuli most similar to the stimulus presented on the preceding trial. This sequential effect is consistent with path-independent models. They also report a second, smaller sequential effect: a tendency *not* to emit those responses that correspond to stimuli most similar to the stimuli presented two to six trials earlier. This effect, which Ward and Lockhead call a contrast effect, is inconsistent with the most obvious generalizations of the simple path-independent models that have been suggested. However, careful examination of the data presented by Ward and Lockhead (1971) indicates that

Fig. 3. Proportion of predictions of a run continuation, Y_k , as a function of run length for the second 250 trials.

this contrast effect is very small and of doubtful significance, except in those conditions in which Ss had displayed before them at the time they made each response the sequence of stimuli that had been presented previously.

In the second experiment, there was some indication that the auditory recognition recency curve became more positive over trials. The tendency for recency curves to become more positive over trials also has been found in binary choice tasks (Brown & Overall, 1959; Bush & Morlock, 1959; Derks, 1962; Edwards, 1961; Estes, 1957; Friedman et al, 1964; Lindman & Edwards, 1961; Nicks, 1959). One possibility is that in the recognition experiments, some Ss, or at least some Ss on some trials, take the guessing attitude and not the judgmental attitude, but that this tendency drops out over trials. Thus, a better estimate of the operation of the biasing mechanism would be obtained after Ss had considerable experience, although for reasons somewhat different from those proposed by Estes (1964) for prediction tasks.

Estes (1964) suggested that the negative recency effect might result from response tendencies the Ss bring into the experiment, and that these tendencies extinguish with experience in the experimental setting. Anderson (1964) argued that positive recency effects may be trained in instead. Neither explanation accounts for the existence of positive recency effects on the early trials of our recognition experiments. It is not so much what Ss come in with; rather, it is what they are asked to do that is important. On the other hand, there is no explanation for why the guessing attitude should drop out over trials.

REFERENCES

- Anderson, N. H. An evaluation of stimulus sampling theory: Comments on Professor Estes' paper. In A. W. Melton (Ed.), *Categories of human learning*. New York: Academic Press, 1964. Pp. 129-144.
- Atkinson, R. C. A variable sensitivity theory of signal detection. *Psychological Review*, 1963, 70, 91-106.

- Atkinson, R. C., & Kinchla, R. A. A learning model for forced-choice detection experiments. *British Journal of Mathematical & Statistical Psychology*, 1965, 18, 183-206.
- Brown, W. L., & Overall, J. E. Implications of recency effects for probability learning theories. *Journal of General Psychology*, 1959, 61, 243-251.
- Bush, R. R., Luce, R. D., & Rose, R. M. Learning models for psychophysics. In R. C. Atkinson (Ed.), *Studies in mathematical psychology*. Stanford: Stanford University Press, 1964. Pp. 201-217.
- Bush, R. R., & Morlock, H. C. Test of a general conditioning axiom for human two-choice experiments. Memorandum MP-1, Department of Psychology, University of Pennsylvania, 1959.
- Derks, P. L. The generality of the "conditioning axiom" in human binary prediction. *Journal of Experimental Psychology*, 1962, 63, 538-545.
- Edwards, W. Probability learning in 1000 trials. *Journal of Experimental Psychology*, 1961, 62, 385-394.
- Estes, W. K. Of models and men. *American Psychologist*, 1957, 12, 609-617.
- Estes, W. K. Probability learning. In A. W. Melton (Ed.), *Categories of human learning*. New York: Academic Press, 1964. Pp. 89-128.
- Friedman, M. P., Burke, C. J., Cole, M., Keller, L., Millward, R. B., & Estes, W. K. Two-choice behavior under extended training with shifting probabilities of reinforcement. In R. C. Atkinson (Ed.), *Studies in mathematical psychology*. Stanford: Stanford University Press, 1964.
- Friedman, M. P., Carterette, E. C., Nakatani, L., & Ahumada, A. Comparisons of some learning models for response bias in signal detection. *Perception & Psychophysics*, 1968, 3, 5-11.
- Jones, M. R. From probability learning to sequential processing: A critical review. *Psychological Bulletin*, 1971, 76, 153-185.
- Lindman, H., & Edwards, W. Supplementary report: Unlearning the gambler's fallacy. *Journal of Experimental Psychology*, 1961, 62, 630.
- Nicks, D. C. Prediction of sequential two-choice decisions from run events. *Journal of Experimental Psychology*, 1958, 57, 105-114.
- Schoeffler, M. S. Theory for psychophysical learning. *Journal of the Acoustical Society of America*, 1965, 37, 1124-1133.
- Taylor, H. A. The observing response in a cued psychophysical task. Technical Report No. 105, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1966.
- Ward, L. M., & Lockhead, G. R. Sequential effects and memory in category judgments. *Journal of Experimental Psychology*, 1970, 84, 27-34.
- Ward, L. M., & Lockhead, G. R. Response system processes in absolute judgment. *Perception & Psychophysics*, 1971, 9, 73-78.

NOTE

1. The data presented in Fig. 1 are only from those Ss who reported that they were still trying to notice stimulus differences and were not trying to discover a pattern in the stimulus presentation schedule at the end of the 350 trials. Three Ss reported otherwise and were discarded. They were replaced by three additional Ss. In the second experiment (Figs. 2 and 3), the data from all Ss that were tested were used.

(Received for publication April 9, 1973;
accepted July 12, 1973.)