

Base-rate training without case cues reduces base-rate neglect

DAVID A. CASE and EDMUND FANTINO
University of California, San Diego, La Jolla, California

and

ADAM S. GOODIE
University of Georgia, Athens, Georgia

Base-rate neglect is a persistent phenomenon in which subjects do not place sufficient weight on the probabilities of occurrence of relevant events. Two experiments with college students support the hypothesis that base-rate neglect may be minimized by providing base-rate training in the absence of case, or witness, cues, prior to introducing (or reintroducing) these cues. In Experiment 1, the hypothesis was supported by both within-subjects and between-groups assessments; in Experiment 2, the hypothesis was supported while the effects of instructions and a correction procedure were found to be minimal. In Experiment 1, but not in Experiment 2, training with case cues present also reduced base-rate neglect, but this effect was not sufficient to account for the effect of cue-absent base-rate training. Correction trials led some subjects to detect that the task contingencies were random; however, neither this nor actually telling subjects after the experiment that the task was indeed random led invariably to subjects' describing the optimal strategy (which was to choose the richer alternative exclusively).

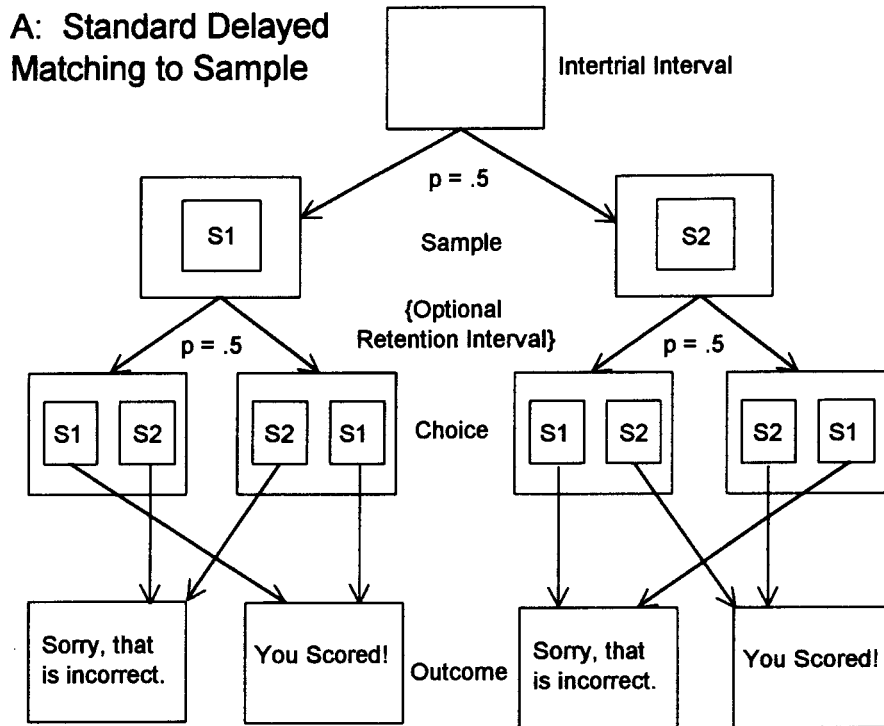
Decisions are sometimes made without fully incorporating relevant information. An important example of information neglect, or selective data utilization, consists of choices in which there is competition between base-rate information and information provided by case, or individuating, cues, also called *witness cues*. Tversky and Kahneman (1982) studied an example in what has become a paradigmatic method. Subjects were asked to rate the likelihood that a cab of a given color, green or blue, was involved in a nighttime crash after being given two items of information: (1) the base rates of occurrence of cabs of those types and (2) an eyewitness's reliability in identifying cabs like the one attested to have been involved, measured under the conditions present at the time of the collision. The different sources of information were put into conflict by Tversky and Kahneman so that one type of cab was implicated by the base rate (e.g., 85% of the cabs in the city were green, 15% blue) and the other type by the witness testimony (i.e., the witness reported seeing a blue cab—the minority color—although the witness was correct only 80% of the time in reliability tests). Subjects' modal likelihood rating of crash involvement ap-

proximately equalled the reliability of the witness (an 80% chance that the cab was blue). The correct probability estimate, given by Bayes's rule, equals a 59% chance that the cab was green, not blue (see Koehlers, 1996, for an explanation of the formula and a recent comprehensive review of these phenomena). More importantly, the subjects' ratings imply that the base-rate information was effectively ignored. This kind of selective attention to or processing of pertinent information is called *base-rate neglect*.

In Goodie and Fantino (1995), instead of employing a verbal task that relied on assumptions about subjects' linguistic backgrounds, subjects repeatedly chose between two alternatives that were preceded by a stimulus. Delayed matching-to-sample (DMTS) was the paradigm used. In a DMTS task, subjects choose among response options associated with stimuli after first seeing one of the stimuli by itself as a sample. Choosing a green alternative after seeing a green sample, for instance, is *matching the sample*. In a common version of the task, this choice would be reinforced every time, and any other alternative would never be reinforced. The stimulus serving as the sample varies from trial to trial, as do the sides of the alternatives (Figure 1A).

Goodie and Fantino (1995) changed the reinforcement contingencies in typical DMTS so that choosing the matching comparison stimulus was reinforced probabilistically. Matching was reinforced on only a percentage of trials, and nonmatching was reinforced on the remaining trials (Castellan, 1977; Estes & Burke, 1955; Humphreys, 1939; Myers, 1976). Goodie and Fantino (1995) then drew the following two parallels to the cab problem of Tversky and Kahneman (1982). First, the unconditional probabilities

The present research was supported by Grant MH57127-1 from the National Institute of Mental Health and Grant IBN94-07206 from the National Science Foundation to the University of California, San Diego, Edmund Fantino principal investigator. We gratefully acknowledge Erasmo Garcia and Will Connor for assistance in collecting the data. Correspondence concerning this article may be addressed to D. A. Case, University of California, San Diego, Department of Psychology, 0109, 9500 Gilman Drive, La Jolla, CA 92093-0109 (e-mail: dcase@ucsd.edu).



Probability of Reinforcement for a Given Choice Following a Given Sample:

	Standard DMTS		Present Experiments	
	Sample		Sample	
	S1	S2	S1	S2
Choice S1	1.0	0	.67	.67
Choice S2	0	1.0	.33	.33

B: Parallels between Questionnaire and Behavioral Task

Testimony of Witness (reliability described)	Samples in DMTS (reliability equals mean of cells in negative diagonal of reinforcement probability matrix)
Base Rates of Cabs (proportion described)	Relative Frequency of Reinforcement of each Choice (frequencies are means of cells in each row of matrix)

Figure 1. (A) Flowchart of standard delayed matching-to-sample task, and matrices of conditional reinforcement probabilities for it and the similar choice task in the present experiments. The stimuli, S1 and S2, were green and blue rectangles displayed on a computer monitor and were counterbalanced across subjects. A 2-sec retention interval was correlated with a blank screen (not shown), the same display as that for the 5-sec intertrial interval. Wording of the outcome screen differed slightly from that which is shown and also displayed the cumulative score. (B) Corresponding features of the Tversky and Kahneman (1982) question about witness to a cab involved in an accident and the present behavioral analogue for studying base-rate neglect.

of reinforcement for choosing a green and for choosing a blue stimulus corresponded to the base rates of green and blue cabs. That is, the unconditional probabilities of reinforcement, regardless of the sample presented, con-

stituted the relevant base rates. Second, the sample stimulus paralleled the witness, or case, cue. The reliability of the sample in predicting which alternative would be reinforced if chosen was less than perfect, just as the wit-

ness in the cab problem did not always correctly distinguish the color of cabs at night (Figure 1B; see Shanks, 1991, for another comparison of experienced versus described probabilistic contingencies).

Goodie and Fantino (1995) found that subjects made nonoptimal choices in the task, neglecting base-rate information as did subjects in the verbal task of Tversky and Kahneman (1982). In a second experiment, substantial incentives were introduced to supplement feedback for correct choices, but these did not alleviate the error significantly. Goodie and Fantino (1996) replicated the basic finding in three further experiments aimed at isolating determinants of base-rate neglect. In these studies, they found that preexisting associations between samples and choice stimuli interfered with optimal choice. However, even when the relation between samples and choice stimuli was totally arbitrary (e.g., vertical line sample positively but imperfectly correlated with reinforcement for choosing the green choice alternative), and even though the samples had no predictive function (i.e., their reliability was 50%), the richer alternative was still chosen considerably less than 100% of the time, as would have been optimal.

In the present experiments, we further pursued the problem of how to reduce base-rate neglect, using the same nonverbal methods that Goodie and Fantino (1995, 1996) used. The main purpose of the present experiments was to assess whether subjects would become more sensitive to base rates if first exposed to a procedure in which base rates need not compete with samples, training that is referred to in other contexts as *probability learning*. Such a result would follow from a parallel to divided attention in perceptual experience.

EXPERIMENT 1

To test this possibility, in the first experiment we studied two groups of college students who responded in the same DMTS procedure as that used previously. In one group (S-NS-S), the subjects were studied in two sessions of DMTS, and then in two sessions with an identical procedure except that the trials did not include a sample. The subjects then returned to the DMTS task (i.e., returned to baseline in an ABA design). This constituted a within-subjects test of the hypothesis that base-rate neglect would be reduced in the second phase of DMTS. For the other group (NS-S), training in the no-samples procedure was given first in the experiment, followed by DMTS. Thus, by comparing the NS-S group's performance in DMTS with that of the S-NS-S group in their first DMTS phase, we made a between-groups assessment of the effect of no-sample training on subsequent DMTS performance. In order to increase sensitivity to base rates, the samples were not predictive of outcomes (i.e., their average reliability was 50%). Although they did not duplicate Tversky and Kahneman's (1982) witness reliability (80%), Goodie

and Fantino (1995, 1996) routinely employed unreliable samples in their successful replications of base-rate neglect; it makes the finding all the more striking.

Two main patterns of data in DMTS suggest that base rates are neglected. The more direct one is the degree to which the minority schedule (i.e., leaner) sample is matched. To see this, first recall that the average sample reliability in the task is 50% combined over samples. Matching the majority schedule sample every time plus countermatching the minority schedule sample every time is an optimal strategy. A subject also could optimize by simply ignoring the case cue, whatever it is, and choosing the majority schedule stimulus every time. Another sign of base-rate neglect is a similar degree of matching of the majority and minority cues, at whatever degree of sample matching there is. We report both measures and also the overall extent of optimal choice combined over case cues. The latter dependent variable is the only one that permits a comparison of performance across sample and no-sample phases (i.e., across DMTS and probability learning tasks).

Method

Subjects and Materials. The subjects were 18 male and female undergraduate psychology students at the University of California, San Diego. Each received credit for 2 h of research participation toward lower division psychology course requirements. The subjects also received monetary payment of \$5 per hour for sessions after the first two, plus \$.01 per point earned in all sessions. Before starting, a sheet detailing the rights of experimental subjects was presented to each individual and consent to participate was obtained. No one declined or withdrew. After completing all aspects of their participation, the subjects were debriefed, paid, and dismissed with directions not to discuss the experiment with others.

The experimental room resembled a typical office with desks, chairs, chalkboard, computer, and books, but no clock. The room had no windows other than a mirrored observation window in the door. An OEI Electronics 386 Turbo personal computer with VGA graphics and a color monitor presented the stimuli, controlled experimental events, and recorded the data. The subjects responded by pressing keys on the computer keyboard.

Procedure. The subjects were first prepared for the task with written instructions, and then they were left alone in the room to repeatedly choose between two side-by-side green and blue rectangles displayed on the computer screen. They made their choices by pressing either the "d" or the "k" key to pick the left or right figure, respectively. These letters were displayed under the rectangles as a repeated reminder of which keys to use. The general instructions indicated how to make choices, that each correct choice would receive a point and what a point was worth monetarily, and that the objective was to earn as much as possible. Also indicated was that in some of the sessions a cue would be presented shortly before the choice display ("your choice will be preceded by a screen containing either a blue or green rectangle."). When a single choice response was made, the rectangles were immediately replaced by an outcome display. This display indicated whether the choice was correct or incorrect, as well as the cumulative number of points in that session. The screen also directed subjects to press any key to continue. When this occurred, the screen darkened for a 5-sec intertrial interval (ITI). Each session comprised 200 trials. In the procedure involving prechoice (i.e., case) cues, a larger rectangle, either green

or blue with equal probability on each trial, was displayed in the center of the screen after the ITI. Below it was the instruction to press any key to continue. Pressing a key darkened the display for 2 sec, and then the choice display was presented. In the procedure without case cues, the choice display was presented immediately after the ITI. The unconditional probability of reinforcement for choosing one color (the majority alternative) was .67, and for choosing the other (the minority alternative), .33; these probabilities were in effect with and without case cues and for either type of case cue (which had no predictive function). The green and blue colors were counterbalanced across subjects and the left-right positions of the colored figures were equiprobable on each trial. All probabilities were pseudorandomly drawn without replacement from sets of 60 elements.

Designs and scheduling of sessions. There were 10 subjects in one group, studied in three successive conditions in an ABA design. The remaining 8 subjects made up a second group, studied in the corresponding BA design. The subjects were randomly assigned to groups. Each A condition consisted of two consecutive sessions of the procedure with case, or sample cues (S), whereas each B condition consisted of two consecutive sessions of the procedure without these cues (NS). One hour was allocated for each session, although subjects always finished considerably sooner. A subject could choose to participate in consecutive sessions on the same day with a short break in between, but usually each session was scheduled on a different day, according to the subjects' convenience.

Results and Discussion

The results from the first condition of the S-NS-S group should show whether base-rate neglect was replicated under the same conditions in which it has been reported elsewhere. This appears to have occurred. For instance, one indication of base-rate neglect is the matching of the minority case cue (instead of countermatching it). With the same probability values and nearly identical conditions and instructions, matching of the minority cue in five prior experiments ranged from 44% to 56% on the average over these first two sessions (two experiments in both Goodie, 1997, and Goodie & Fantino, 1995; and one experiment in Goodie & Fantino, 1996). In the present case, the minority cue was matched 54% of the time. The effect of sessions was nonsignificant in this phase.

Another gauge of base-rate neglect comprises similar levels of matching of the two cues. The difference in matching percentage between the majority and minority stimuli ranged from 5% to 20% across the studies of Goodie and Fantino, whereas in the present study this difference was 11%.

Finally, recall that the optimal strategy is to choose the majority alternative exclusively, regardless of the cue presented prior to choice. Group S-NS-S subjects in their first condition chose the richer outcome on only 56% of all trials, producing a relatively large loss of potential rewards. This percentage is closer to the average cue accuracy (50%) than to optimal performance (100% selection of the richer outcome).

The most sensitive measure of a predicted reduction in base-rate neglect resulting from training without case cues is the difference in DMTS between the initial and final conditions in the S-NS-S group. A $2 \times 2 \times 2$ within-

subjects analysis of variance was conducted to assess this, with the factors before/after phase, majority/minority cue, and first/second session within each phase. Matching-to-sample was significantly greater with the majority cue than with the minority cue, 66% versus 46% [$F(1,9) = 12, p < .01$; note that in all analyses we employed $p < .05$ as our criterion for significance, but we also report $p < .01$ as is conventional; all statistics are rounded to two significant digits]. Greater matching of the majority cue was exhibited by 9 of the 10 subjects.

The main effect of greatest interest—the effect of training without case cues in the intervening NS phase—also was significant [$F(1,9) = 10$]. Following training in which case cues were withheld, average matching-to-sample across cues declined to 53% from 59% before this training. Crucially, the direction of change depended on the type of cue: for the two-way interaction, $F(1,9) = 5.4$ (see Figure 2A). Thus, as predicted, matching the majority cue increased after prior training without cues, while matching the minority cue decreased. The increased matching of the majority cue was small on the average and occurred in just half of the subjects. By contrast, the critical measure of reduced base-rate neglect—decreased matching of the minority cue—was substantial and occurred in 9 of the 10 subjects, with an average decline from 54% to 39%. Clearly, by this evidence, subjects came to be more sensitive to base rates as a function of training with base rates in a procedure wherein attention to them was not diverted by case cues.

The remaining main effect of sessions within phases also was significant. There was less matching-to-sample with more training, 59% versus 54% [$F(1,9) = 22, p < .01$]. The fact that performance changed across training within a phase raises the question of the extent to which effect of training without case cues is attributable to that training. That is, the effect could be a result of time in the experiment, regardless of the type of training. The following between-group contrasts for assessing the effect of training without case cues address this issue.

A between-groups test is generally less sensitive than a within-groups test, but has the advantage of being less dependent on contestable statistical assumptions needed for repeated measures experimental designs (see, e.g., Kepple, 1973, pp. 394–400). In this case, comparing the two groups permits an assessment of whether reduced base-rate neglect is a function of specific training without case cues versus time in the experiment—that is, training regardless of presence or absence of case cues. We compare first the matching performance of the NS-S group with that of the S-NS-S group in its first DMTS condition. One difference between these groups is that the NS-S group had prior training without case cues (NS phase), whereas the S-NS-S group did not have any at this point. As in the within-groups assessment, the main effect of cue type was reliable and large. The majority cue was matched on 70% of trials on the average over

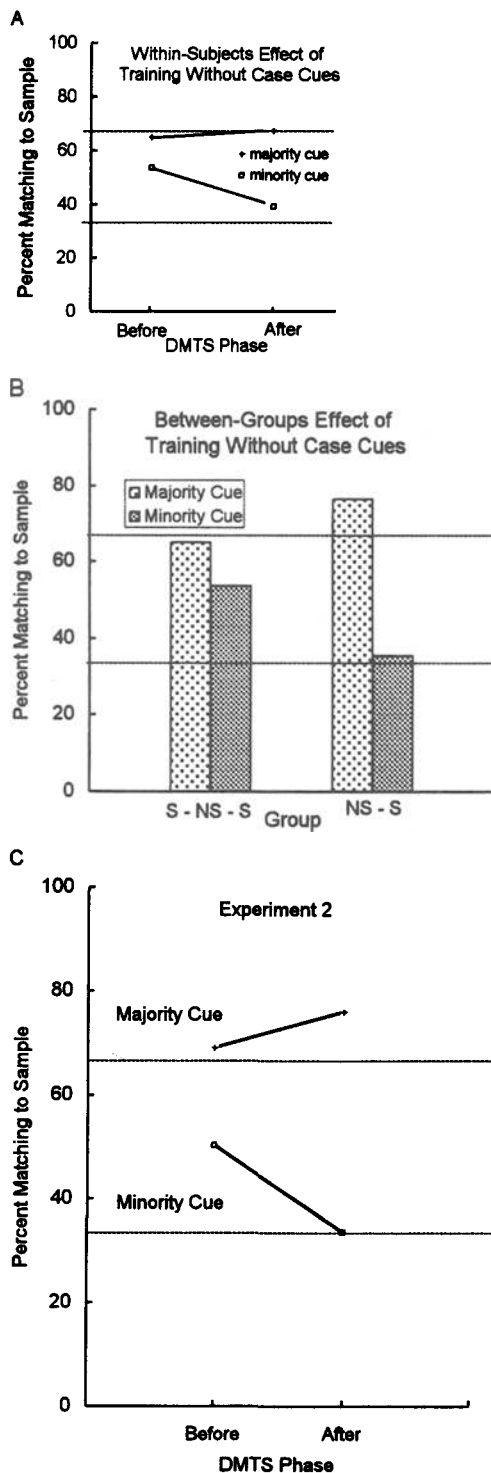


Figure 2. (A) Mean percentage of times that subjects in the within-subjects assessment group matched the sample cues in each phase of DMTS in Experiment 1. In the training phase intervening between these phases, choices of the same alternatives were reinforced at the same base rates but no samples were presented. The horizontal lines are the base rates of reinforcement for choosing the cues. (B) Mean percentage of times that subjects in the two groups in Experiment 1 matched the sample cues in the first (Group S-NS-S) or only phase (Group NS-S) of DMTS. (C) Mean percentage of times that subjects in Experiment 2 matched the sample cues in each phase of DMTS.

groups and sessions, while the minority cue was matched 46% of the time [$F(1,16) = 21, p < .01$]. The effect of type of cue also interacted with the difference between groups, as it did in the within-subjects assessment (Figure 2B). Although the majority cue was always matched more than the minority cue, this difference in matching between cues was greater in Group NS-S, which had had prior training without samples cues. Specifically, prior training resulted in moderately increased matching of the majority cue but substantially decreased matching of the minority cue. Thus, the between-groups evidence suggests again that base-rate training decreased base-rate neglect.

This is not to say that training duration was not implicated. In fact, increased sensitivity to base rates was also a function of length of training within a phase [for the sample by session interaction [$F(1,16) = 4.5$]. Matching the majority stimulus increased slightly from 68% to 72% across the two sessions, while matching the minority stimulus decreased to a somewhat greater degree, from 49% to 42%. Since the three-way interaction was not significant, these between-groups data join the within-groups contrast above in suggesting that training per se may reduce base-rate neglect quite apart from training without case cues present. Note that, even if correct, this does not imply that base-rate training per se is ineffective. It merely implies that, in the analyses up to this point, the data have not eliminated the possibility that training per se—with case cues or without them—is critical to reduced base-rate neglect.

If length of training was indeed critical to reduction in base-rate neglect regardless of whether training involved case cues or not, then base-rate neglect ought to be significantly less in the final S phase of the S-NS-S group than in the S phase of the NS-S group, because the S-NS-S group had twice as much time in the experiment than did the NS-S group at the beginning of the final phase. This hypothesis was not borne out, however. In this analysis, the effect of cue type was robust and large as usual [majority vs. minority cue, 71% vs. 37%, $F(1,16) = 23, p < .01$], but the F ratio for the interaction of group and cue type was less than 1. That is, subjects matched the respective cues essentially identically in both groups despite the extra training in the S-NS-S group by this point in the experiment.

In the same analysis, length of training within a phase interacted with cue type [$F(1,16) = 6.3$]. Specifically, matching the majority cue increased minimally across sessions (71% vs. 72%) while matching the minority cue clearly decreased (41% vs. 33%). This is the same kind of within-phase training effect as that which appeared in two analyses above. Since no other interactions were significant, the conclusion is that training with case cues has the same kind of effect as does training in the absence of case cues; namely, both types of training can reduce base-rate neglect. However, when the two types of training were combined in the final DMTS phase of the S-NS-S group, the effect of training per se did not add to the effect of NS training, as can be seen by contrasting

these results with those of the NS-S group. Had the S-NS-S group shown less base-rate neglect (and thus more optimal responding) than the NS-S group did in this comparison, the results would have called into question whether there is a special advantage to removing distractors—the case cues—in learning to attend to base-rate information.

Given that training without the interference of case cues improves sensitivity to base rates, choice of the richer schedule across all trials may be less when case cues are present than when they are absent. To test this, we compared the groups' average performance across phases in the NS-to-S transition. If training without case cues eliminates base-rate neglect, there should be little difference between performances in these two phases. Indeed, there were no significant differences, which suggests that case-cue-induced base-rate neglect had been completely eliminated in the final DMTS phase. However, the richer schedule was still chosen only 66% of the time, not 100% as would be optimal. We will return to this discrepancy in the General Discussion.

EXPERIMENT 2

In Experiment 2, we replicated the conditions of the S-NS-S group from the first experiment and examined the effect of two new manipulations. Using a 2×2 between-groups factorial design, we instructed half the subjects about the nature of the task at the beginning of the no-samples phase. Our intent was to improve performance by conveying with simple "picture instructions" the random nature of correct responses and the exact base rate that had been and would again be experienced. Note that instruction procedures may be viewed as intermediate between the experiential "trial and error" methods of Goodie and Fantino (1995, 1996) and the verbal hypothetical scenarios of Tversky and Kahneman (1982). The data will enable us to assess what role verbal aspects have beyond those of task contingencies.

With the second manipulation, we inquired whether a correction procedure in the no-samples phase would similarly facilitate attention to base-rate information. In this procedure, when a given alternative was randomly determined to be correct by the computer program on a (noncorrection) trial, that alternative had to be selected by subjects before a new correct alternative would be programmed. Thus, incorrect choices were followed by correction trials. This implies that in the correction procedure, regardless of subjects' choices, each subject necessarily experienced the exact base rates of reinforcement programmed for each alternative in each session. This additional "feedback" and greater precision of information might be expected to enhance performance. However, note that the correction procedure inevitably also requires that subjects choose the minority alternative more often than would be optimal in the standard

noncorrection procedure. For this reason and because such choices are reinforced, it is possible that correction training in the no-samples phase—despite the connotation of edification implied in its name and the enhanced task information provided by the correction contingency—might instead impair subsequent DMTS performance (even though there correction trials are no longer programmed). In either event, effects of instructions and correction training on subsequent DMTS can be compared with those provided by training without case cues present but where performance is neither corrected nor instructed as described above.

Method

Subjects and Materials. The subjects were 16 male and female undergraduate psychology students at the University of California, San Diego, who were recruited and compensated in the same manner as were those in Experiment 1. The materials and scheduling of sessions were the same as in Experiment 1. A brief paper-and-pencil questionnaire was administered immediately after the experiment.

Procedure. All the subjects were studied in the same ABA design as in Experiment 1, except that these subjects were randomly divided into four groups that differed in terms of their treatment in the NS phase. One group was treated the same as the S-NS-S group in Experiment 1. Half the subjects had instructions given to them just prior to the beginning of the NS phase, which constituted one factor in the 2×2 factorial design; the other between-groups factor was whether a correction procedure was used in the no-samples phase. The manipulated written instructions read in part:

In the next phase of the experiment the choice screens will not be preceded by a blue or green rectangle. The array of letters shown below illustrate a batch of typical trials in the procedure (B and G stand for the blue and green alternatives). While almost certainly not identical to what you will experience, this sequence was generated in exactly the same way that the computer will generate the sequence of correct alternatives in your sessions.

The subjects were then required to accurately count the 67 letters of the majority outcome and 33 letters of the minority outcome before beginning the first session of the NS phase.

A correction procedure was in effect in the NS phase for half the subjects. First, a correct outcome for a trial was randomly selected by the computer as usual. If a subject chose that outcome, the trial was identical to the noncorrection procedure (and to that of Experiment 1). Choice of the other outcome also produced the same screen message as before ("That is incorrect," etc.), but in the correction procedure the alternative that was correct remained the same on subsequent correction trials, as many as were needed until it was finally chosen. In other words, the other outcome could not be selected randomly by the computer to be correct until the prior choice had been corrected. To the subjects, the correction trials appeared to be identical to the original choice trials, including the random alternation of sides on which the alternatives appeared.

Results and Discussion

The minority cue was matched 50% of the time on the average in each session of the first DMTS phase, which is precisely the reliability of the samples. Also, the difference in matching percentage between the majority and minority stimuli averaged only 19% in this phase. Each of these indicators of base-rate neglect falls within the range of data of prior studies as described above. Sub-

jects chose the majority stimulus 59% of the time on the average over both sample cues in these sessions, a level nearly equal to that in Experiment 1 (56%) and again far from the optimal strategy (100%). Nonoptimal choice and base-rate neglect are robust phenomena under the conditions that have been studied.

A five-factor mixed-design analysis of variance was conducted to assess whether training without case cues reduced base-rate neglect. The within-subjects design that was used in Experiment 1 was augmented here by two between-groups factors, which were (1) whether or not subjects received correction trials in the no-samples phase and (2) whether or not subjects were instructed before beginning the no-samples phase.

The sole significant main effect on proportion matching-to-sample was whether the sample was the majority or minority cue [$F(1,15) = 30, p < .01$] (see Figure 2C). And only one interaction was significant, that of cue type (i.e., majority vs. minority) \times phase of DMTS training [$F(1,12) = 7.0$]. As is shown in Figure 2C, matching the majority cue increased in the second DMTS phase compared with the first, while matching the minority cue decreased, as in Experiment 1. Thus, choice became more optimal and more sensitive to base rates as a function of prior training without case cues. These effects were obtained regardless of instructions or a correction procedure in the no-samples phase. Such robustness of results despite method variation can be taken as support for the generality of the benefits of training without case cues.

In every other respect except one, the performance of subjects in the group studied in an exact replication of Experiment 1 were comparable to subjects' performance in that study. The exception was that in Experiment 2 there was no effect of sessions of training within phases. Although the explanation of this discrepancy between experiments is unclear, the lack of an effect of sessions makes it ever more evident that training without case cues in the no-samples phase was instrumental to the enhanced performance in the DMTS task in both experiments. Moreover, the results of Experiment 2 suggest that simply giving further training in the DMTS task does not itself necessarily lead to reduction in base-rate neglect. The issue of how to better—or ideally, totally—eliminate base-rate neglect is addressed in the General Discussion.

Failure to find the expected effect of instructions cannot be attributed to an inadequately large group size, since performance with instructions was actually worse on the average than performance without them. Also, DMTS performance was slightly worse as a function of correction training. The group that did best in Experiment 2 was therefore the one receiving neither instructions nor correction trials in the no-samples phase (i.e., the subjects exposed to exactly the same methods as in Experiment 1). Although the groups were not significantly different from each other in the first DMTS phase,

a test of effects of instructions and the correction procedure based on individual before-and-after difference scores was also conducted; means of the four groups remained nonsignificantly different in this analysis.

Answers to the postexperimental questionnaire revealed that correction training in the no-samples phase was not totally without effect, however. In one questionnaire item, subjects were asked to choose between two answers: Were correct choices in the task random? Or, were they based on a rule or pattern? Among subjects who had received correction trials, 5 guessed that events were random, whereas none did so who had been exposed to the standard probability learning task in the phase intervening between the DMTS assessments ($\chi^2 = 7.3, p < .01$).

In a subsequent question on another sheet of paper, subjects were first told quite explicitly that correct choices in the experiment were in fact randomly determined. Subjects were also told the base rates. Then they were asked to "describe how you would perform to maximize earnings if you could continue participating for one more session." Only half indicated in their answers the maximization strategy of exclusively choosing the majority cue. Interestingly, these answers did not associate significantly with the answers to the earlier question, nor were these groups' prior mean DMTS performances significantly different in the final phase ($p > .4$). In other words, subjects who had previously correctly answered that correct choices in the experimental task were randomly determined were no more likely to describe correctly how to maximize earnings than those who had previously incorrectly indicated that a rule or pattern determined correct DMTS choices. In agreement with this finding, mean DMTS performance of the 5 aforementioned subjects from the two correction groups also was not significantly greater than the mean of the remaining subjects in the final phase ($p > .2$). Thus, correction training in the no-samples phase had an effect on certain relevant questionnaire responses but not on others or on performance in the DMTS task. The experimenter literally telling subjects that the task was random, even after subjects had correctly concluded on the basis of their experience that it was random, did not then lead automatically to a description of optimal performance upon request. This constitutes a striking instance of information neglect and nonoptimal reasoning that should be able to facilitate theoretical analysis.

GENERAL DISCUSSION

In two experiments on DMTS performance, case cues and base rates were placed into competition for control of choice. The experiments tested whether experience with the base rates in the absence of the case cues would reduce base-rate neglect as assessed when case cues were introduced or reintroduced. Support for this possi-

bility was found in both within-subjects and between-groups comparisons in Experiment 1, and in the demonstration of generality across two method variations in Experiment 2—instructions and correction training—that plausibly might have modulated the outcome. These instructions did not use the word *random*, nor did they verbally quantify the base rates using an explicit description of the percentages. Therefore, they cannot be viewed as a command to perform in a certain way, which would merely test compliance to an experimenter-provided rule, rather than effects of the cognitive aspects of the situation. Neglect of base rates did not decline by simply providing additional training in DMTS in Experiment 2, but in Experiment 1 additional DMTS training appeared to be effective.

Half the subjects in Experiment 2 failed to describe the optimal response strategy after the experiment ended, even though they were told then that correct choices in the task were determined randomly. The subjects who articulated that to maximize earnings one must choose the majority stimulus every time did not actually choose the majority cue significantly more than the less astute subjects. This dissociation of verbal (i.e., questionnaire) and nonverbal (i.e., task) behavior is in agreement with the lack of effect of the instructions that were manipulated in the task. Correction training in the no-sample phase led some subjects to posit on their own that correct choices were randomly determined. However, those who correctly deduced this about the task did not choose the majority cue reliably more than the subjects who indicated that correct choices followed a pattern or rule.

The degree to which subjects choose the richer schedule more than the leaner schedule is a direct indication of their approximation to optimizing monetary gain. At the same time, if preference for the richer schedule when samples are absent equals that when samples are present—as was ultimately true in the present study—then base-rate neglect may seem an inappropriate description of the present performance in DMTS. Base-rate neglect conventionally means merely that when explicitly manipulated case cues and base rates compete for control of behavior, and Bayes's theorem is the referee of the contest, base rates come up short. Although subjects eventually learned to attend to base rates equally whether case cues were present or not, they still did not utilize base-rate information optimally, suggesting the need for a revised definition. At any rate, a contributing factor to this nonoptimality is that subjects may generate "hypotheses" for predicting correct responses, and then these hypotheses may compete with the optimal "random rule" for control of choice (i.e., "choose the richer alternative exclusively, because with randomly determined outcomes, that is the most successful strategy"). If these incorrect implicit guesses could be undermined, as the samples were eliminated in the no-samples phase of the present experiments, perhaps subjects would choose the richer schedule exclusively;

that is, they would choose optimally. One method of undermining subjects' self-generated hypotheses may involve actual experience in a random task during which systematic variation in performance is induced by extra-task instructions or contingencies. In this manner, the relation between choice and deviations from optimal performance could be detected by the subject. Fictitious self-generated hypotheses would be dissociated from covariation of (extraneously manipulated) choices and degree of success. If such a method would work in one situation, additional experiences like this across a variety of situations would likely suffice to generally eliminate base-rate neglect.

A disadvantage of exclusively choosing the richer alternative is the sacrifice of information about what might have happened had a subject continued to sample the leaner alternative. That is, we have defined optimal choice within the rather narrow confines of a particular, simple task for which the base rates do not vary across time. In a different situation in which base rates changed across time, exclusive choice of the currently richer alternative might not be optimal, depending on the frequency and extent of changing contingencies. It may be that the subjects in Experiment 2 who indicated that they would continue to respond to the minority cue even after being told the nature of the task were sensitive to the advantage of continuing to sample alternatives in varying environments, because that may in fact be closer to the rule for extralaboratory contingencies than situations in which base rates are fixed for long periods. In any event, subjects in the present experiments persisted in base-rate neglect over hundreds of repetitive trials, a tendency that was sometimes lessened with sufficient training and that was minimized by training without potentially competing information from case cues. The results suggest that base-rate neglect is a robust phenomenon, occurring in a behavioral task similar to how it does in questionnaires (e.g., Goodie & Fantino, 1995, 1996; Stolarz-Fantino & Fantino, 1990, 1995), but that the degree of neglect may be reduced with training that allows undivided attention to base-rate information.

REFERENCES

- CASTELLAN, N. J., JR. (1977). Decision making with multiple probabilistic cues. In N. J. Castellan, Jr., D. P. Pisoni, & G. R. Potts (Eds.), *Cognitive theory* (Vol. 2, pp. 117-147). Hillsdale, NJ: Erlbaum.
- ESTES, W. K., & BURKE, C. J. (1955). Application of a statistical model to simple discrimination learning in human subjects. *Journal of Experimental Psychology*, *50*, 81-88.
- GOODIE, A. S. (1997). Base-rate neglect under direct experience. *Dissertation Abstracts International*, *58*, 435B.
- GOODIE, A. S., & FANTINO, E. (1995). An experientially derived base-rate error in humans. *Psychological Science*, *6*, 101-106.
- GOODIE, A. S., & FANTINO, E. (1996). Learning to commit or avoid the base-rate error. *Nature*, *380*, 247-249.
- HUMPHREYS, L. G. (1939). Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *Journal of Experimental Psychology*, *25*, 294-301.

- KEPPEL, G. (1973). *Design and analysis: A researcher's handbook*. Englewood Cliffs, NJ: Prentice-Hall.
- KOEHLERS, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral & Brain Sciences*, **19**, 1-53.
- MYERS, J. L. (1976). Probability learning and sequence learning. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 3, pp. 171-205). Hillsdale, NJ: Erlbaum.
- SHANKS, D. R. (1991). On similarities between causal judgements in experienced and described situations. *Psychological Science*, **2**, 341-350.
- STOLARZ-FANTINO, S., & FANTINO, E. (1990). Cognition and behavior analysis: A review of Rachlin's "Judgment, decision, and choice." *Journal of the Experimental Analysis of Behavior*, **54**, 317-322.
- STOLARZ-FANTINO, S., & FANTINO, E. (1995). The experimental analysis of reasoning: A review of Gilovich's "How we know what isn't so." *Journal of the Experimental Analysis of Behavior*, **64**, 111-116.
- TVERSKY, A., & KAHNEMAN, D. (1982). Evidential impact of base rates. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 153-160). Cambridge: Cambridge University Press.

(Manuscript received September 11, 1997;
revision accepted for publication September 14, 1998.)