The Relation between Order Effects and Frequency Learning in Tactical Decision Making

Jiajie Zhang[1], Todd R. Johnson[2], & Hongbin Wang[1]
Department of Psychology[1]
Division of Medical Informatics[2]
The Ohio State University
Columbus, Ohio 43210
{zhang.52, johnson.25, wang.190}@osu.edu

ABSTRACT

This article presents three experiments that examine the relation between order effects and frequency learning, with the following results. First, when frequencies of occurrence are presented as sequences of real events, base rates can be learned and used with a high degree of accuracy. However, conditional probabilities for multiple sequentially presented evidence items cannot be completely learned, due to the distortion of a recency order effect for actual decisions. Second, there is also a recency order effect for belief evaluations, which cannot be eliminated even if base rates are used correctly. Third, base rates learned in one environment can be transferred to another environment, but the transfer soon diminishes due to learning in the new environment. However, belief evaluations are not transferred from one to another environment. The existing models of frequency learning cannot explain the order effect for actual decisions because they do not consider sequential information. The existing models of belief updating can explain both types of order effects, but they do not have any mechanisms for frequency learning. To account for the complete spectrum of frequency learning and order effects, we outline our initial effort in developing a unified model that integrates frequency learning and order effects.

Belief updating and frequency learning are two ubiquitous phenomena. Suppose that a disease R can be diagnosed by test A and test B. When a physician knows the result of test A, the physician's belief of disease R is at certain level. When the physician subsequently gets the result of test B, the physician's previous belief of disease R based on test A will be updated. This is the phenomenon of belief updating, that is, the updating of belief values for sequentially presented information items. Let us further suppose that the results of test A and test B are associated with disease R with certain frequencies. With many diagnoses of disease R based on test A and test B, the physician may acquire some knowledge about these frequencies. This is the phenomenon of frequency learning. Although belief updating and frequency learning often coexist in many tasks, as in the diagnosis task just described, their studies are usually separated.

For belief updating, a number of studies indicate that the order in which evidence items are presented can affect the strength of a person's belief in hypothesized causes. For example, the physician's final belief value of disease R after receiving the test results in A-B order might be different from the final belief value after receiving the same test results in B-A order. This is called the order effect, which is a very robust effect in belief updating tasks (for a review, see Hogarth and Einhorn, 1992). Different experiments have found that order of evidence items can produce a recency effect, a primacy effect, or no effect. A recency effect is that the final belief value is mainly determined by the last evidence item in the sequence, a primacy effect is that the final belief value is mainly determined by the initial evidence item in the sequence, and no-effect means that the final belief evaluation is independent of the order of evidence items. Hogarth and Einhorn (1992) proposed an anchoring and adjustment model of belief updating that predicts when these effects will occur based on whether evidence is encoded in evaluation or estimation mode and whether the belief updating process is step-by-step or end-of-sequence. In evaluation tasks, people encode each evidence item as positive (confirms) or negative (disconfirms) relative to a hypothesis, whereas in estimation tasks, people assess a moving average that reflects the po-

This research was supported by Grants N00014-95-1-0241 and N00014-96-1-0472 from the Office of Naval Research, Cognitive and Neural Sciences & Technology Division, and by a summer fellowship from the Center for Cognitive Science at The Ohio State University. Part of the current work was presented at the Eighteenth Annual Conference of the Cognitive Science Society. Correspondence and request for reprints should be addressed to Jiajie Zhang, Department of Psychology, or Todd R. Johnson, Department of Pathology, The Ohio State University, Columbus, Ohio 43210, USA.

sition of each new evidence item relative to the current hypothesis. The step-by-step process requires subjects to express their beliefs after integrating each evidence item in a given sequence, whereas the end-of-sequence process requires subjects to report their beliefs only after all evidence items have been presented. As an example of predications, the anchoring and adjustment model predicts that step-by-step evaluation of beliefs for mixed positive and negative evidence items produces a recency effect, whereas step-by-step evaluation of beliefs for consistent evidence items (all positive or all negative) produces no effect. For the disease example, if we assume that the diagnosis is a step-by-step evaluation task and that test A is positive and test B is negative, then the physician's conclusion will be biased by the evidence presented last (a recency effect). The physician might conclude that disease R does not exist if positive test A is presented first and negative test B is presented second. If the test order is reversed, the physician might conclude that R is present.

The frequency literature is mostly concerned with the learning and use of frequency information. When frequencies of occurrence are presented in terms of probabilities or percentages, they are very difficult to use, often resulting in various biases such as the base rate fallacy, conjunction fallacy, representativeness heuristic, overconfidence, and so forth (e.g., Bar-Hillel, 1980; Casscells, Schoenberger, & Graboys, 1978; Lyon & Slovic, 1976; Kahneman, Slovic & Tversky, 1982). However, when frequencies of occurrence are presented in frequency formats, the rate of using them correctly can be dramatically increased (e.g., Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995). Studies of frequency learning show that when frequencies are presented in terms of sequences of real events and occurrences (sometimes called natural sampling, see Kleiter, 1994), they can often be learned automatically and used with a high degree of accuracy (e.g., Carroll & Siegler, 1977; Christensen-Szalanski, & Beach, 1982; Christensen-Szalanski, & Bushyhead, 1981; Manis, Dovalina, Avis, & Cardoze, 1980; for a review, see Hasher & Zacks, 1984). As a result of using sequences of real events to present frequency information, many of the biases disappear. For example, the base-rate fallacy, which is the neglect of base rate information, was even exhibited by trained physicians when base rate information was presented in terms of probabilities in pencil-and-paper problems (Casscells, Schoenberger, & Graboys, 1978). But when base rate information was obtained from experience in terms of learning from sequences of real events, the base-rate fallacy disappeared (Christensen-Szalanski, & Bushyhead, 1981).

Due to the lack of studies that jointly consider frequency learning and belief updating, models of frequency learning cannot account for frequency learning for sequentially presented information items, and models of belief updating cannot account for belief updating that involves frequency learning. Since many frequency learning tasks involve belief updating for temporal sequences of information, as in the diagnosis task described above, it is important to study the relation between frequency learning and belief updating. For example, it is important to know whether frequency learning and belief updating are independent processes. It would also be interesting to see whether the order effect, which is a robust reasoning bias, would disappear when frequencies of occurrence are acquired accurately from sequences of real events.

The objective of this article is to examine the relation between belief updating and frequency learning in a common task environment. Specifically, we design three experiments to study the relation between the order effect and the learning of base rates and conditional probabilities. Experiment 1 examines whether base rates and conditional probabilities can be acquired automatically and correctly from sequences of real events and whether there is an order effect during and after frequency learning. Experiment 2 examines the same issues of Experiment 1 with different, non-neutral base rates. Experiment 3 examines whether the frequency information learned from one environment can be transferred to a new environment and whether the belief evaluations in one environment can be transferred to a new environment. Based on the experimental results, we will evaluate the existing models of frequency learning and belief updating and outline a preliminary model that integrates frequency learning and belief updating.

EXPERIMENT 1

The experimental task was implemented in the CIC (Combat Information Center) simulator developed by Towne (1995) for the US Navy. Figure 1 shows an oversimplified schematic drawing of a portion of the CIC display. It shows an unknown airplane, called a contact, heading towards the Navy ship that is at the center of the radar display. The captain of the ship can check whether the contact is on or off a commercial air route by clicking the route button to display all available routes. The routes are immediately available after clicking the route button. The captain can also send a radio verbal warning (MAD 1, or Military Air Distress Level 1) to request the contact to identify itself by clicking the warning button. The contact may or may not respond to the warning. In the three experiments reported here, when it responds to the warning, it always identifies itself as a commercial airplane. If there is a response, it comes back about 30 seconds after the warning is issued. The contact can be either friendly or hostile. The task is to use the information about the air route and the information about the identity (ID) obtained from the radio warning to identify whether the contact is friendly or hostile. The constraint of the task is that the two evidence items (route and ID) can only be obtained sequentially, one at a time. The order can be either route followed by ID or ID followed by route.

This experiment examines three issues. The first is about frequency learning. In a given geopolitical environment, there are certain base rates for friendly and hostile contacts and conditional probabilities of friendliness or hostility given the two evidence items. Previous studies on frequency learning show that base rates and conditional probabilities can be learned automatically with a high degree of accuracy when the evidence items are presented in parallel (e.g., Gluck & Bower, 1988, 1990; Shanks, 1990a, 1990b; for a review, see Hasher & Zacks, 1984). The current experiment examines whether base rates and conditional probabilities can be learned accurately for sequentially presented evidence items. The second issue is the order effect for actual decisions during frequency learning. Because the evidence items are presented sequentially, the order of presentation may affect how the decisions are made. This is a phenomenon that has not been previously studied. The third issue is the order effect for belief evaluations. Previous studies show that when frequency information is accurately and automatically learned in real events, certain biases such as the base rate fallacy can be eliminated. The current experiment examines how the acquired base rates and conditional probabilities affect the order effect (a special type of bias) for belief evaluations after frequency learning.

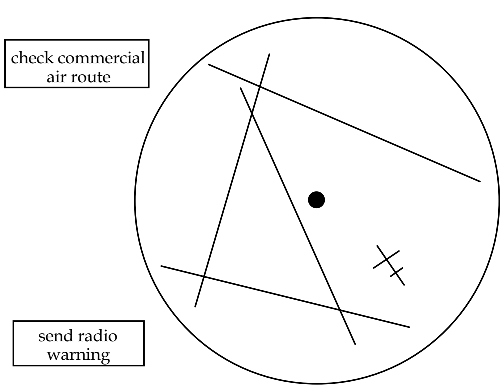

Figure 1. Simplified radar display on the CIC simulator. See text for details.

Table 1. Probability Distributions of Learning Trials in the Experiments

		Exp 1	Exp 2 & 3	
	Total # of Contacts	50	75	75
	Base Rates	F:H = 1:1	F:H = 2:1	F:H = 1:2
Given Information	$p(R+\mid F)$	0.80	0.80	0.80
	$p(R+\mid H)$	0.20	0.20	0.20
	$p(ID+\mid F)$	0.80	0.80	0.80
	$p(ID+\mid H)$	0.20	0.20	0.20
Calculated from Bayes Rule	$p(F\mid R+)$	0.80	0.89	0.67
	$p(F\mid R-)$	0.20	0.33	0.11
	$p(F\mid ID+)$	0.80	0.89	0.67
	$p(F\mid ID-)$	0.20	0.33	0.11
	$p(F\mid R+ID+)$	0.94	0.97	0.89
	$p(F\mid R-ID-)$	0.06	0.11	0.03
	$p(F\mid R+ID-)$	0.50	0.67	0.33
	$p(F\mid R-ID+)$	0.50	0.67	0.33

* F = Friendly, H = Hostile

Method

Subjects. The subjects were 40 undergraduate students in introductory psychology courses at The Ohio State University who participated in the experiment for course credit.

Design & Procedure. There were two independent evidence items: Route and ID. Route indicates whether the contact is on or off a commercial air route. ID indicates whether there is any response from the unknown contact to a radio warning issued from the ship. The two evidence items were presented in two different orders: *Route-ID* in which Route information was collected first, followed by ID information, and *ID-Route* in which ID information was collected first, followed by Route information. After having collected both evidence items, subjects made a forced-choice decision indicating whether the unknown contact was friendly or hostile. After each decision, subjects were given feedback indicating whether the decision was correct or incorrect. Each trial lasted for about one minute. Each subject performed 50 trials for a total of 50 contacts in a different random order. For half of the 40 subjects, the two evidence items were always presented in the Route-ID order for all 50 trials; for the other half, in the ID-Route order for all 50 trials. The 50 trials for each subject constituted the learning phase for the acquisition of frequency information. It took about one hour to finish the 50 trials. Among the 50 contacts in the 50 trials, 25 were friendly and 25 were hostile, that is, the base rate is F:H = 1:1 (F = Friendly, H = Hostile). The conditional probabilities were designed as follows. If a contact was friendly, there was an 80% chance that it was on a commercial air route and an

80% chance that it gave a positive (commercial) ID; and if a contact was hostile, there was a 20% chance that it was on a commercial air route and 20% chance that it gave a positive ID. From these values, the conditional probabilities of friendliness and hostility for a given set of evidence items were calculated from Bayes rule, as shown in Table 1. The following are two examples of the calculations.

$$p(F \mid R+) = \frac{p(F)p(R+ \mid F)}{p(F)p(R+ \mid F) + p(H)p(R+ \mid H)} = \frac{0.50 \times 0.80}{0.50 \times 0.80 + 0.50 \times 0.20} = 0.80$$

$$p(F \mid R + ID+) = \frac{p(F)p(R+ \mid F)p(ID+ \mid F)}{p(F)p(R+ \mid F)p(ID+ \mid F) + p(H)p(R+ \mid H)p(ID+ \mid H)}$$

$$= \frac{0.50 \times 0.80 \times 0.80}{0.50 \times 0.80 \times 0.80 + 0.50 \times 0.20 \times 0.20} = 0.94$$

After 50 learning trials, each subject was given a written questionnaire requesting belief evaluations about the hostility and friendliness of an unknown contact after the presentation of a baseline fact and each of the two evidence items. The written questionnaire constituted the evaluation phase for belief evaluations. In the questionnaire, Route was always negative (the plane was not on a commercial route) and ID was always positive (the plane indicated that it was a commercial plane). For half of the 20 subjects receiving each of the two learning orders (Route-ID and ID-Route), the evaluation order of the evidence items was Route-ID; and for the other half, the evaluation order was ID-Route. Thus, this experiment was a 2×2 between-subject design with 10 subjects in each of the four conditions, with the two learning orders as one factor and the two evaluation orders as another factor. As an example, the questionnaire for the Route-ID evaluation order is as follows.

1. *You see a plane which is getting closer to your ship. On a scale from 0 to 100 (with 0 being total disbelief and 100 total belief) please rate your belief in the following hypotheses:*
    1. *How likely do you think the plane is hostile?*
    2. *How likely do you think the plane is friendly?*

2. *After consulting commercial air routes, you discover that the plane is not on a commercial air route. Given this new information, please answer the following questions. Again, express your answer on a scale of 0 to 100 with 0 being total disbelief and 100 being total belief.*
    1. *How likely do you think the plane is friendly?*
    2. *How likely do you think the plane is hostile?*

3. *When you asked the plane to identify itself, the plane identifies itself as a commercial airplane. Given this new information, please answer the following questions. Again, express your answer on a scale of 0 to 100 with 0 being total disbelief and 100 being total belief.*
    1. *How likely do you think the plane is friendly?*
    2. *How likely do you think the plane is hostile?*

Results

Frequency Learning. The decisions for the 50 trials by each subject were transformed into observed base rates and conditional probabilities of friendliness given a set of evidence items, which were then averaged across the 20 subjects for the Route-ID learning order and across the 20 subjects for the ID-Route learning order. The results are shown in Figure 2 under the labels *Route-ID* and *ID-Route*. Note that the base rates and conditional probabilities calculated from the 50 trials are not the values after subjects had finished learning. Instead, they are the values that included the initial trials at the beginning stage of the learning. Thus, they may not reflect the true values that subjects had learned. The reason we did not divide the 50 trials into an early learning set and a late learning set and then calculate conditional probabilities separately is because 50 trials are the minimum set of trials that could generate the given frequency distribution.

For both the Route-ID and ID-Route learning orders, the observed base rates were not significantly different from the normative value ($t(19) = 0.782$, $p = 0.44$, and $t(19) = 0.424$, $p = 0.68$, respectively). They did not differ from each other, either ($t(19) = 0.238$, $p = 0.81$).

The observed conditional probabilities from the experiment were compared with their corresponding normative values, which were calculated from the frequency values given in the experimental design by Bayes rule (see Table 1). For the Route-ID learning order, p(F|ID+), p(F|ID-), and p(F|R-ID-) were significantly different from the normative values (smallest $|t(19)| = 2.82$, $p < 0.05$); the other five conditional probabilities were not significantly different from the normative values (largest $|t(19)| = 2.0$, $p > 0.05$). For the ID-Route learning order, p(F|ID+), p(F|ID-), p(F|R-ID-), p(F|R+ID-), and p(F|R-ID+) were significantly different from their normative values (smallest $|t(19)| = 2.26$, $p < 0.05$); the other three conditional probabilities were not significantly different from the normative values (largest $|t(19)| = 1.31$, $p = 0.20$). These results indicate four findings. First, even if the calculations were based on the 50 trials that included the initial imperfect trials, subjects correctly learned some of the conditional probabilities. Second, subjects learned more for the Route-ID order than for the ID-Route order. Third, p(F|ID+) was smaller than its normative value but p(F|R+) was not different from its normative value, indicating that ID might be a less objective evidence item than Route. This might be because Route was always objectively observed from the radar display whereas ID was obtained from possibly deceptive radio communication. Fourth, p(F|ID-) was larger than its normative value but p(F|R-) was not different from its normative value, indicating that a negative ID suggested less hostility than a negative Route. This might be because a negative ID is subject to multiple interpretations, such as a friendly contact with malfunctioned radio or with a functional radio which is not turned on, or a friendly contact whose pilot did not hear the message or did not understand the language. Although a negative Route is also subject to multiple interpretations, such as a commercial airplane that went off course, it seems that subjects did not consider these alternative interpretations as very likely. The incorrect learning of the probabilities that deviated from their normative values might be due to the subjectivity of a positive ID and the multiple interpretations of a negative ID. It is also likely that the deviations were simply a reflection of the imperfect learning for the initial trials, which were included in the calculations of these probabilities.

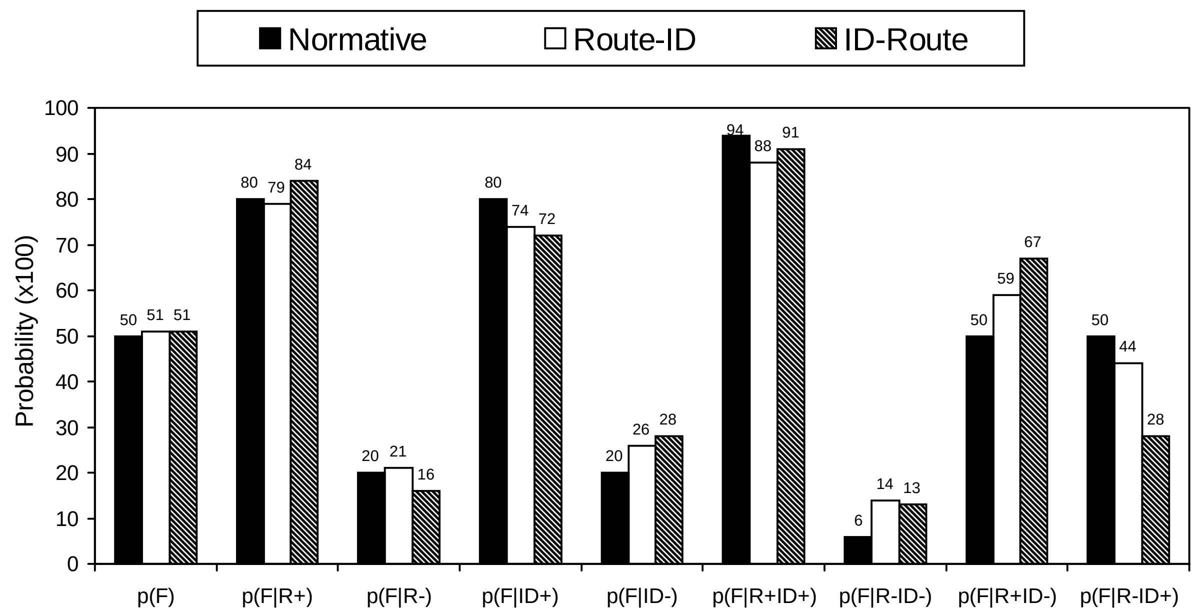

Figure 2. The observed and normative base rates and conditional probabilities for the two learning orders in Experiment 1.

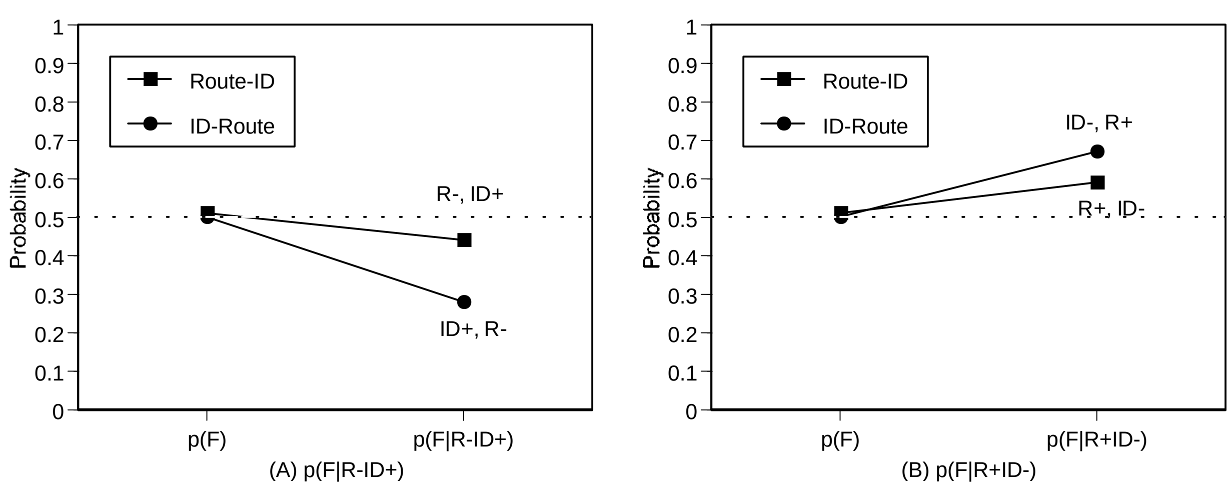

Figure 3. Recency order effect for actual decisions during frequency learning. The dashed lines are the normative base rates.

The conditional probabilities from the two different learning orders were also compared with each other. The only significant difference was p(F|R-ID+) ($F(1, 38) = 4.52$, $p < 0.05$). This is clearly a recency order effect for mixed positive and negative evidence items: p(F|R-ID+) for the Route-ID order was larger than p(F|R-ID+) for the ID-Route order because the last evidence item in the Route-ID order was positive (ID+) whereas that in the ID-Route order was negative (R-). A similar order effect was also observed for p(F|R+ID-) although the difference between the two learning orders was not statistically significant ($F(1, 38) = 0.707$, $p = 0.41$). Figure 3 explicitly shows this result. This result is an indication that subjects showed an order effect for actual decisions during frequency learning. The insignificant differences between the two learning orders for p(F|R+ID+) and for p(F|R-ID-) suggest that the order effect for actual decisions did not appear for consistent (all positive or all negative) evidence items.

Belief Evaluation. The results of belief evaluations after the learning phase are shown in Figure 4. There was a clear order effect: when the two evidence items (positive ID, negative Route) were presented in different temporal orders, the final friendliness evaluations of the unknown contact were different. An ANOVA for the final evaluations of friendliness was conducted for the two learning orders and two evaluation orders. The interaction between learning order and evaluation order was not significant ($F(1, 38) = 0.53$, $p = 0.47$). The main effect of learning order was significant ($F(1, 38) = 4.32$, $p < 0.05$), indicating that the learning order Route-ID produced more hostile evaluations than the learning order ID-Route. The main effect of evaluation order was also significant ($F(1, 38) = 13.41$, $p < 0.001$): the evaluation order Route-ID produced a more friendly final evaluation than the evaluation order ID-Route, indicating an order effect for evaluations. This order effect is a recency effect: the final evaluation of friendliness was determined by the last evidence item. For the Route-ID evaluation order, the last evidence item, ID, was positive, producing a more friendly evaluation. In contrast, for the ID-Route evaluation order, the last evidence, Route, was negative, producing a more hostile evaluation.

To test how well subjects used the base rate information that they acquired during frequency learning, subjects' evaluations at the base level for the two learning orders were compared with the normative evaluation value at the base level, which is 50. There were no significant differences between subjects' evaluations and the normative value for either learning order (largest $|t(19)| = 2.03$, $p > 0.05$). This indicates that subjects appeared to have used the base rate correctly.

The effect of learning order on evaluations was also analyzed as a mixed design for each evaluation order with the two learning orders as the between-subject factor and the three evaluations (base, datum1, and datum2) within each evaluation order as the within-subject factor. For the Route-ID evaluation order, the interaction between learning orders and evaluations was not significant ($F(2, 36) = 0.333$, $p = 0.72$). The main effect of learning order was marginally significant ($F(1, 18) = 2.94$, $p = 0.10$), indicating that the learning order Route-ID produced slightly more hostile evaluations. The main effect of individual evaluations was significant ($F(2, 36) = 6.59$, $p < 0.005$), indicating that belief values changed when new evidence items became available. For the ID-Route evaluation order, the results were the same.

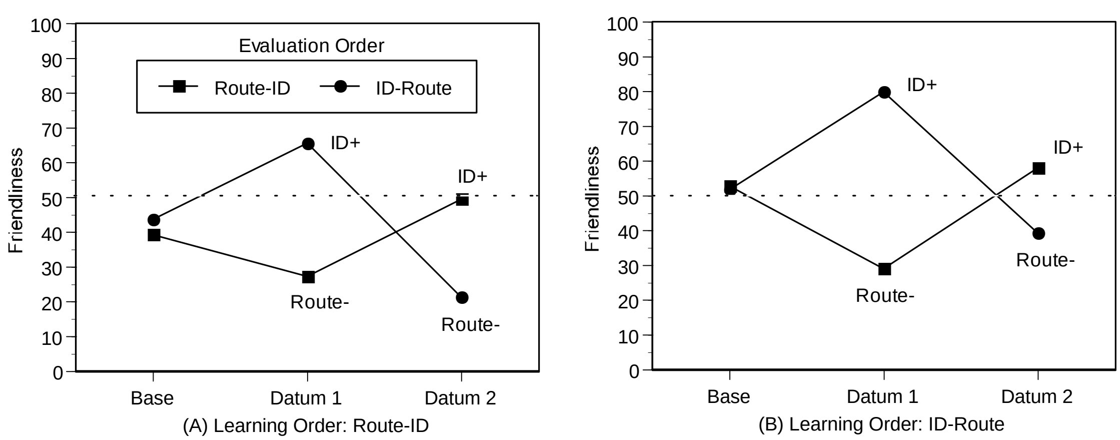

Figure 4. Order effect for the evaluations of friendliness after frequency learning. The dotted lines represent the normative values for base rate F:H = 1:1.

The interaction was not significant ($F(2, 36) = 1.30$, $p = 0.28$). The main effect of learning order was significant ($F(1, 18) = 10.95$, $p < 0.005$), again indicating that the learning order Route-ID produced more hostile evaluations than the learning order ID-Route. The main effect of evaluations was also significant ($F(2, 36) = 35.33$, $p < 0.001$), again indicating that belief values changed when new evidence items became available.

Discussion

This experiment generated the following results. First, subjects showed a recency order effect for actual decisions during frequency learning for mixed positive and negative evidence items, which has not been reported previously. The decision task in the learning phase of the current experiment appeared to be an end-of-sequence task in which subjects made a decision after having observed both evidence items. If it was indeed an end-of-sequence task, then it can not be explained by Hogarth and Einhorn's (1992) anchoring and adjustment model, which would predict that for an end-of-sequence evaluation task with mixed evidence items, the order effect should be a primacy effect. However, if we assume that subjects behaved in a step-by-step manner with an implicit belief evaluation after each evidence item was presented, the recency order effect is exactly what Hogarth & Einhorn's model predicts. This assumption of implicit belief evaluations was likely to be true because the 30 second interval between the first and second evidence items in each learning trial was long enough for subjects to do some belief evaluations.

Second, when subjects acquired frequency information from sequences of real events, they accurately acquired the base rates. In addition, subjects also correctly used the base rate information they learned for belief evaluations. However, the correct learning and use of base rate information in this experiment might be merely an apparent phenomenon because the 50% base rates for both friendly and hostile contacts would be expected even if subjects ignored base rates. Experiment 2 was designed to test this possibility.

Third, the recency order effect for mixed evidence items for actual decisions has important implications for the study of frequency learning. For example, p(F|R-ID+) for the Route-ID learning order was larger than p(F|R-ID+) for the ID-Route learning order even if they should have the same normative value. This is a recency effect because the last evidence item in the Route-ID order was positive (ID+) whereas that in the ID-Route order was negative (R-). Due to this order effect, the conditional probabilities for sequentially presented multiple evidence items can not be completely learned. For this reason, any models of frequency learning, if they are proposed to account for the full spectrum of frequency learning phenomena, should consider the temporal order in which evidence items are presented. The Rescorla-Wagner model (Rescorla & Wagner, 1972; see also Gluck & Bower, 1988; Shanks, 1990b), one of the well-known models that predicts accurate acquisition of frequency information, completely ignores the temporal sequence of information. It should be modified to take into account temporal information.

Fourth, subjects showed a recency order effect for step-by-step belief evaluations on mixed positive and negative evidence items after frequency learning. This is the order effect directly predicted from Hogarth and Einhorn's (1992) anchoring and adjustment model. Since subjects correctly learned and used the base rate in this case, it suggests that the correct learning and use of base rate information could not eliminate the order effect.

EXPERIMENT 2

Subjects correctly learned and used the base rate information in Experiment 1. However, it might be merely an apparent phenomenon because the 50% base rates for both friendly and hostile contacts would be expected even if subjects ignored base rates. Experiment 2 was designed to test whether subjects could indeed correctly learn and use base rate information. There were two base rate conditions: F1-H2 in which the ratio of friendly and hostile contacts is 1:2 and F2-H1 in which the ratio is 2:1. If subjects could indeed learn base rates correctly, then they should learn these two non-neutral base rates because if they ignore base rates, their performance should reflect the 50% neutral base rate.

Method

Subjects. The subjects were 40 undergraduate students in introductory psychology courses at The Ohio State University who participated in the experiment for course credit.

Design & Procedure. The procedure was identical to that in Experiment 1. The design was similar to that in Experiment 1 except that the base rates were different and Route-ID was the only learning order. Each subject performed 75 trials for a total of 75 contacts in a different random order. For half of the 40 subjects, the contacts were friendly in 50 trials and hostile in 25 trials, that is, the base rate was F:H = 2:1 (the *F2-H1* condition). For the other half, the base rate was reversed, which was F:H = 1:2 (the *F1-H2* condition). The conditional probabilities of hostility and friendliness for a given set of evidence items are shown in Table 1. The two evidence items, Route and ID, were independent. The 75 trials for each subject constituted the learning phase for the acquisition of frequency information. After 75 trials, subjects were given the same written questionnaires as in Experiment 1 for belief evaluations. For half of the 20 subjects receiving each of the two different base rates, the evaluation order was Route-ID. For the other half, the order was ID-Route. Thus, this experiment was a 2x2 between-subject design with 10 subjects in each condition, with the two base rates as one factor and the two evaluation orders as the other factor.

Results

Frequency Learning. Similar to Experiment 1, the responses of the 75 trials by each subject were transformed into observed base rates and conditional probabilities, which were then averaged across the 20 subjects in each base rate condition. The results are shown Figure 5.

The observed base rate for the F2-H1 condition (62%) was significantly smaller than its normative value 67% ($t(19) = -2.47$, $p = 0.02$) but larger than the chance value 50% ($t(19) = 6.65$, $p < 0.001$). In contrast, the observed base rate for the F1-H2 condition (39%) was significantly larger than the corresponding normative value 33% ($t(19) = 2.57$, $p = 0.02$) but smaller than the chance value 50% ($t(19) = -5.33$, $p < 0.001$). This result indicates that subjects acquired the base rates in the correct direction but did not acquire the numeric values perfectly. This might be because the calculations of the observed base rates included the initial learning trials that might not be perfect.

The observed conditional probabilities from the experiment were compared with their corresponding normative values calculated from Bayes rule. For the F2-H1 base rate condition, p(F|ID+), p(F|R+ID+), and p(F|R-ID+) were all significantly smaller than their corresponding normative values (smallest $|t(19)| = 4.036$, $p < 0.05$); the other five conditional probabilities were not significantly different from their corresponding normative values (largest $|t(19)| = 0.023$, $p = 0.98$). For the F1-H2 base rate condition, p(F|ID-), p(F|R-ID-), and p(F|R+ID-) were all significantly larger than their corresponding normative values (smallest $|t(19)| = 3.406$, $p < 0.05$); the other five conditional probabilities were not significantly different from their corresponding normative values (largest $|t(19)| = 0.073$, $p = 0.94$). These results indicate two findings. First, subjects correctly learned most of the conditional probabilities, even if the base rates were not 50% vs. 50% for friendly and hostile contacts. Second, once again, ID was a less reliable evidence item than Route. In a friendlier environment (F2-H1 condition), ID+ was a less powerful positive evidence item: the conditional probabilities involving ID+ were all smaller than their normative values. In a more hostile environment (F1-H2 condition), ID- was a less powerful negative evidence item: the conditional probabilities involving ID- were all larger than their normative values.

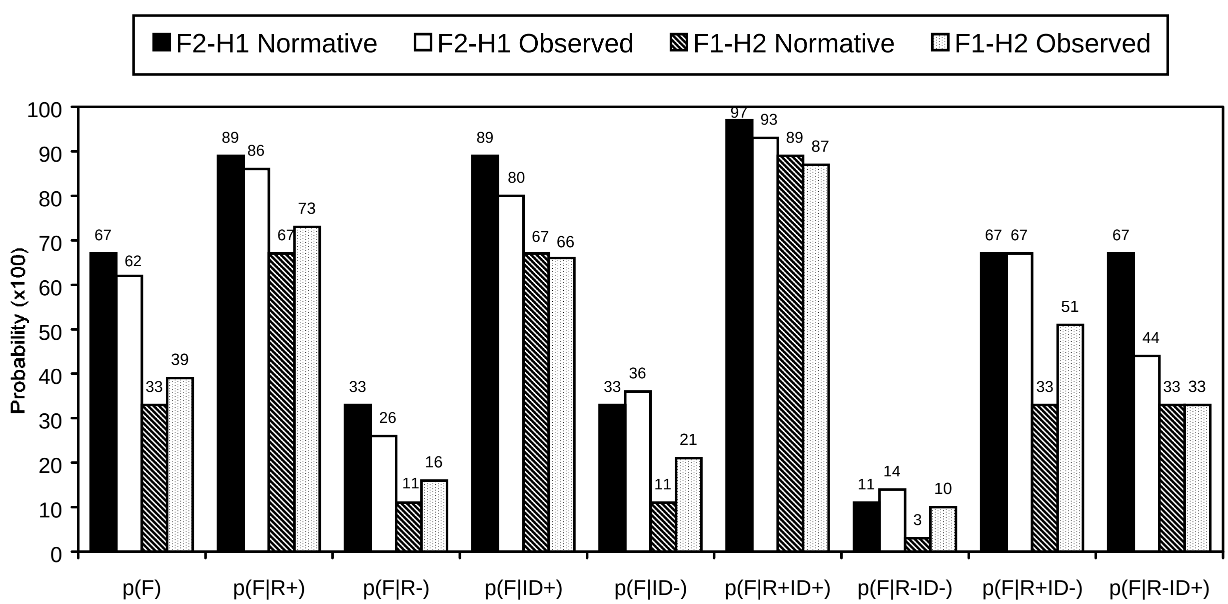

Figure 5. Observed and normative base rates and conditional Probabilities for the two base rate conditions.

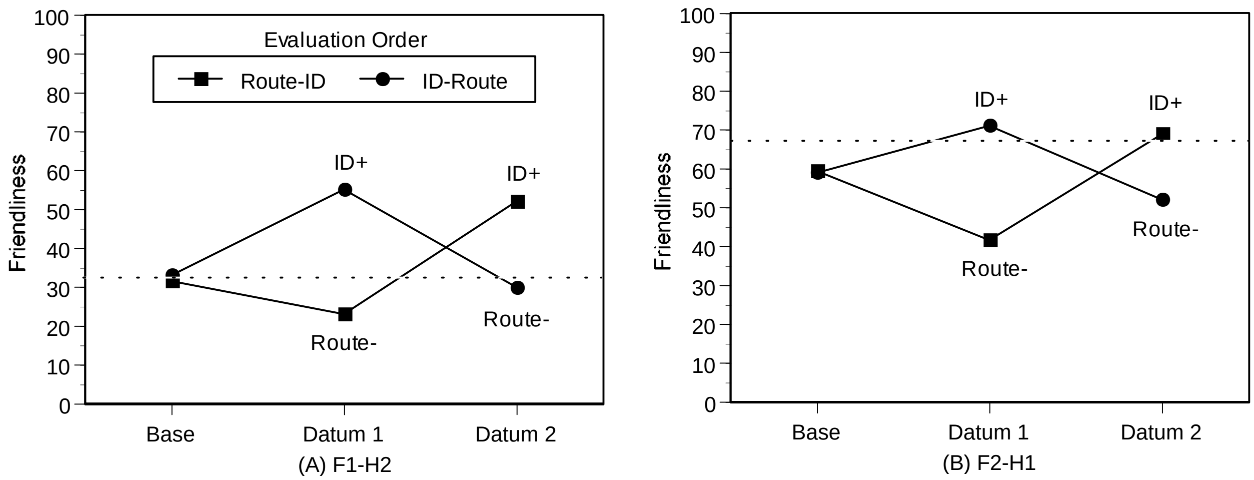

Figure 6. The evaluations of friendliness for the two evaluation orders for the two base rate conditions. The dotted lines represent the normative values for base rates, F:H = 1:2 and F:H = 2:1, respectively.

Belief Evaluation. The results of belief evaluations after the learning phase are shown in Figure 6. There was a clear order effect: when the two evidence items (positive ID, negative Route) were presented in different temporal orders, the final friendliness evaluations of the unknown contact were different. An ANOVA for the final evaluations of friendliness was conducted for the two base rates and the two evaluation orders. The interaction was not significant ($F(1, 36) = 0.0625$, $p = 0.80$). The main effect of base rates was significant ($F(1,36)=4.416$, $p<0.05$), indicating that subjects in the F2-H1 base rate condition gave higher friendly evaluations than those in the F1-H2 base rate condition. This implies that subjects learned and used the base rate information in the correct direction. The main effect of evaluation orders was also significant ($F(1,36) = 4.461$, $p < 0.05$): the evaluation order Route-ID produced higher friendly evaluations than the ID-Route order. Since ID is always positive and Route is always negative, this order effect was a recency effect: the last evidence item has a greater impact on the final evaluations.

To test the effect of base rates on evaluations, an ANOVA was also conducted for each evaluation order with the two base rates (F1-H2 and F2-H1) as the between-subject factor and the three evaluations (base, datum1, and datum2) within each evaluation order as the within-subject factor. For the Route-ID evaluation order, the interaction was not significant ($F(2, 36) = 0.582$, $p = 0.55$). The main effect of base rates was significant ($F(1, 18) = 7.70$, $p = 0.01$), indicating that a higher base rate of friendly contacts produced more friendly evaluations. The main effect of individual evaluations was also significant ($F(2, 36) = 8.69$, $p < 0.001$), indicating that belief values changed when new evidence items became available. For the ID-Route evaluation order, the results were the same. The interaction was not significant ($F(2, 36) = 0.229$, $p = 0.80$). The main effect of base rates was significant ($F(1, 18) = 13.63$, $p = 0.002$), again indicating that a higher base rate of friendly contacts produced more friendly evaluations. The main effect of evaluations was also significant ($F(2, 36) = 4.83$, $p = 0.01$), again indicating that belief values changed when new evidence items became available.

To further test how well subjects used the base rate information, subjects' evaluations at the base level for the two base rate conditions were compared with the normative evaluation values at the base level (67 and 33 for F2-H1 and F1-H2, respectively). There were no significant differences between subjects' evaluations and the normative values for either condition (largest $|t(19)| = 1.23$, $p = 0.25$). This suggests that that subjects appeared to have used the base rates perfectly.

Discussion

This experiment showed the following results. First, although subjects did not learn the base rates perfectly during actual decisions, they did learn them in the correct direction. This indicates that subjects were sensitive to base rates: they did not neglect them. Subjects also correctly acquired most of the conditional probabilities. The ones they did not correctly acquire were those that depended on ID, which might have distorted the learning because it was a less reliable evidence item. Second, once again, after frequency learning, subjects showed a recency order effect for step-by-step belief evaluation of mixed evidence items, as predicted by Hogarth and Einhorn's (1992) model. Third, the base rates learned by subjects during frequency acquisition were correctly used in belief evaluations, with higher friendly evaluations for the F2-H1 condition and lower friendly evaluations for the F1-H2 condition. In addition, it appeared that subjects' use of base information in belief evaluations after learning was nearly perfect: the observed and normative values of evaluations at the base level did not differ. However, the correct use of base rates did not eliminate the order effect for belief evaluations, indicating that the learning and use of base rates did not interact with the order effects of evaluations.

EXPERIMENT 3

Experiments 1 and 2 showed that subjects learned the base rates with a high degree of accuracy and used the base rates nearly perfectly. Experiment 3 was designed to examine (a) whether the base rates learned in one environment could be transferred to a new environment, and (b) whether the belief evaluations in one base rate condition could be transferred to a new environment.

Method

Subjects. The subjects were 24 undergraduate students in introductory psychology courses at The Ohio State University who participated in the experiment for course credit.

Design & Procedure. The design and procedure were similar to those in Experiments 1 and 2, except for the following changes. For all the trials in the learning phase, the two evidence items were always presented in Route-ID order. Each subject performed 150 trials in two stages, which took about two hours to finish. In the first stage, all 24 subjects performed 75 trials with a different random order for each subject with base rate F:H=2:1. This is the learning phase of the first stage. Then the subjects were given the same questionnaire for the R-ID+ evaluation order as used in Experiments 1 and 2. This is the evaluation phase of the first stage.

In the second stage, all 24 subjects performed another set of 75 trials with a different random order for each subject with base rate F:H=1:2, still in the Route-ID learning order. This is the learning phase of the second stage. Then, half of the 24 subjects were given the questionnaire in R-ID+ order and the other half the questionnaire in ID+R- order. This is the evaluation phase of the second stage.

Results

The analyses were focused on the second stage of the experiment, since the purpose of the first stage was to provide the context for the transfer at the second stage. In fact, the first stage is identical to the F2-H1 condition of Experiment 2. The second stage had the same base rate as the F1-H2 condition of Experiment 2, but the former was preceded by the learning of the F2-H1 base rate information whereas the latter was not preceded by the learning of any base rate information. Thus, if there is any transfer effect, it can be revealed by comparing the second stage of the current experiment with the F1-H2 condition of Experiment 2.

Base Rate Learning. The results of base rate learning are shown in Figure 7. The solid circles represent the data of the current experiment and the open circles represent the data of Experiment 2. The base rate learned at the end of the first stage was compared with the base rate learned at the end of the F2-H1 condition of Experiment 2 (shown by the leftmost data points labeled by F2-H1 on the x-axis). There was no significant difference ($t(42) = 0.16$, $p = 0.87$), indicating that the first stage of the current experiment replicated the result of the F2-H1 condition of Experiment 2. In addition, similar to the F2-H1 condition of Experiment 2, the base rate for the first stage (60%) was significantly smaller than the normative value 67% ($t(23) = -4.19$, $p < 0.001$) but larger than the chance value 50% ($t(23) = 6.66$, $p < 0.001$). This indicates that subjects acquired the base rates in the correct direction but did not acquire the numeric values perfectly.

In order to examine the transfer effect, the base rates for the second stage were calculated cumulatively for initial trials 1-5, 1-10, 1-20, 1-30, and 1-75 and for last trials 71-75, 66-75, 56-75, and 46-75. These cumulative base rates were compared with the corresponding cumulative base rates for the non-transfer F1-H2 condition of Experiment 2. The difference was significant for 1-5, 1-20, 1-30, and 1-75 (smallest $|t(42)| = 2.09$, $p < 0.05$) and marginally significant for 1-10 ($t(42) = 1.87$, $p = 0.07$), but it was not significant for 71-75, 66-75, 56-75, and 46-75 (largest $|t(42)| = 0.32$, $p = 0.75$). These results indicate that the base rate information learned in the F2-H1 environment in the current experiment was transferred to the F1-H2 environment, resulting in more friendly base rates for the initial trials. When subjects performed more and more trials for the new F1-H2 environment, the transfer effect became weaker and weaker and eventually diminished at the end of the 75 trials. Thus, base rate information could be transferred from one environment to another environment, but the transfer effect would be soon overcome by the learning of the base rate information of the new environment.

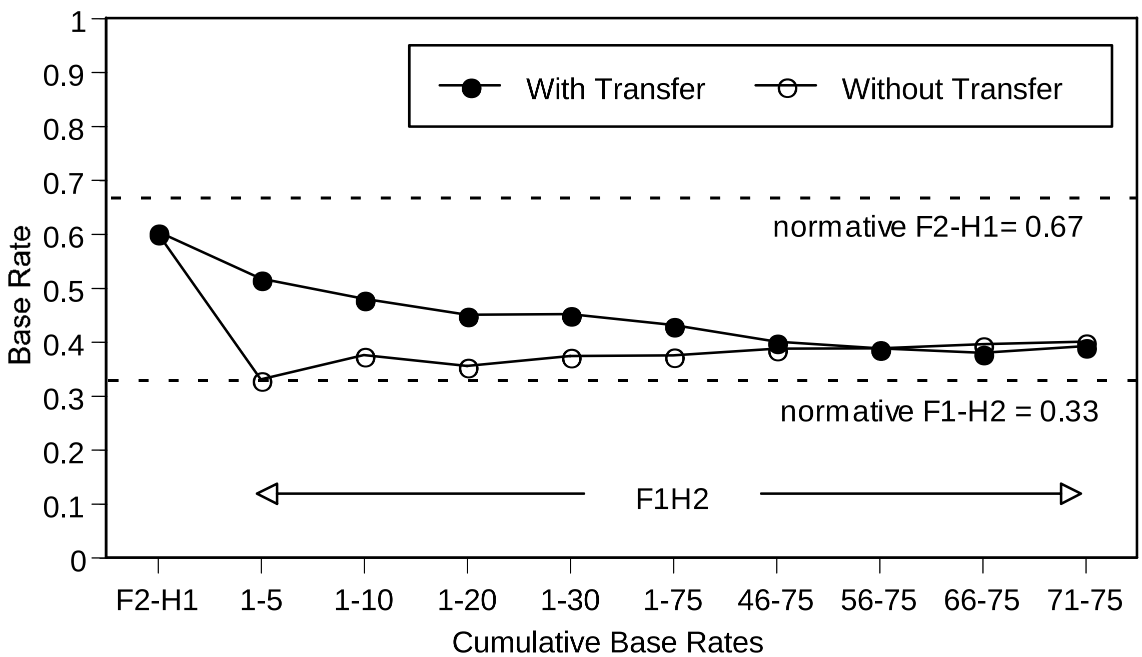

Figure 7. The results of base rate learning. The solid circles represent the data of the current experiment and the open circles represent the data of Experiment 2. F2-H1 on the x-axis refers to the base rate learned in the first stage of the current experiment and that in the F2-H1 condition of Experiment 2. The other values on the x-axis are cumulative base rates learned over a certain number of trials for the second phase of the current experiment and the F1-H2 base rate condition of Experiment 2. For example, 1-5 is the base rate for the initial 5 trials and 66-75 is the base rate for the last 10 trials.

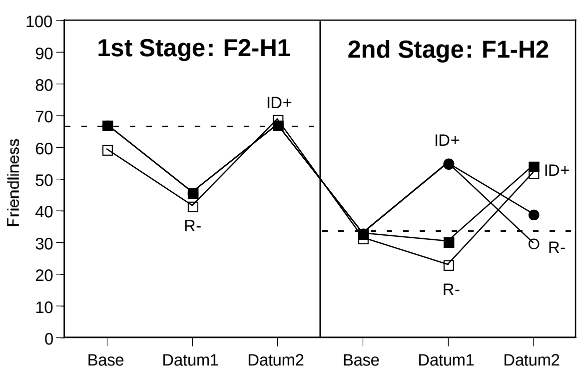

Figure 8. The results of belief evaluations. The solid circles and squares represent the data of the current experiment and the open circles and squares represent the data of Experiment 2. The left panel refers to the first stage of the current experiment compared with the R-ID+ evaluation order of the F2-H1 condition of Experiment 2, and the fight panel refers to the second stage of the current experiment compared with the R-ID+ and ID+R- evaluation orders of the F1-H2 condition of Experiment 2.

Belief Evaluation. The results of belief evaluations are shown in Figure 8. The solid circles and squares represent the data of the current experiment and the open circles and squares represent the data of Experiment 2.

The left panel of Figure 8 shows the evaluations for the R-ID+ evaluation order of the first stage of the current experiment and the evaluations for the corresponding R-ID+ evaluation order of the F2-H1 condition in Experiment 2. A two-way ANOVA for the two experiments and the three evaluations (base, datum1, and datum2) showed that the difference between experiments was not significant ($F(1,20) = 0.007$, $p = 0.94$), nor was the interaction between experiments and evaluations ($F(2,40) = 0.19$, $p = 0.83$). This suggests that the result of Experiment 2 was replicated by the current experiment. In addition, similar to Experiment 2, the base level evaluation value in the current experiment was not significantly different from its normative value ($t(11) = 0.42$, $p = 0.68$), suggesting that the base rate was used correctly.

The right panel of Figure 8 shows the evaluations for the second stage of the current experiment and the corresponding evaluations for the non-transfer F1-H2 condition of Experiment 2. For the R-ID+ evaluation order, a two-way ANOVA for the two experiments and the three evaluations showed that the difference between experiments was not significant ($F(1, 20) = 0.007$, $p = 0.94$), nor was the interaction between experiments and evaluations ($F(2, 40) = 0.19$, $p = 0.83$). For the ID+R- evaluation order, similarly, the difference between experiments was not significant ($F(1, 20) = 0.17$, $p = 0.68$), nor was the interaction between experiments and evaluations ($F(2, 40) = 0.25$, $p = 0.78$). Because the second stage of the current experiment was the transfer condition and the F1-H2 condition of Experiment 2 was the non-transfer condition, the insignificant difference between them suggests that the belief evaluations for the first stage with F2-H1 base rate were not transferred to the second stage with F1-H2 base rate. The order effect for the evaluations in the second stage was in the direction of a recency effect. However, it was not significant: the final evaluation of the R-ID+ evaluation order was not significantly different from the final evaluation of the ID+R- evaluation order ($t(22) = -1.40$, $p = 0.18$). The base level evaluations for the two evaluation orders in the second stage were not significantly different from the normative values ($t(11) = 0.087$, $p = 0.93$ for the R-ID+ order and $t(11) = -0.077$, $p = 0.94$ for the ID+R- order), suggesting that the base rate was used correctly.

Discussion

This experiment showed the following results. First, the base rate learned in one environment was transferred to another environment, but the transfer effect was soon overcome by the learning of the base rate in the new environment. Thus, with sufficient training, the base rate in a new environment could be eventually correctly learned. Second, belief evaluations carried out for one base rate environment were not transferred to a new base rate environment. This result is not surprising because belief evaluations for both environments were performed after the learning was complete. Once the learning was complete, the knowledge of the old environment was completely overridden by the knowledge of the new environment. Third, the first stage of the current experiment was identical to one of the conditions in Experiment 2. The insignificant difference between them indicates a good replication of experimental findings.

GENERAL DISCUSSION

Studies of the order effect are usually separated from studies of frequency learning and use. Since many frequency learning tasks involve multiple sequentially presented information items for each learning trial, it is important to consider the joint implications of these two research areas. The present study explored the relation between order effects and frequency learning in a realistic tactical decision making task, with three major findings.

The first finding is about frequency learning. When frequencies of occurrence are presented in terms of sequences of real events, base rates can be learned and used with a high degree of accuracy. This result converges with similar findings in early studies (Carroll & Siegler, 1977; Christensen-Szalanski, & Beach, 1982; Christensen-Szalanski, & Bushyhead, 1981; Manis, Dovalina, Avis, & Cardoze, 1980; for a review, see Hasher & Zacks, 1984). However, not all frequency information can be correctly learned and used, even when frequencies of occurrence are presented in terms of sequences of real events. Conditional probabilities for multiple sequentially presented evidence items cannot be completely learned, due to the distortion of an order effect for actual decisions. For example, $p(F|R\text{-}ID+)$ learned for the Route-ID learning order was larger than $p(F|R\text{-}ID+)$ learned for the ID-Route learning order even if they should have the same normative value. This is a recency effect because the last evidence item in the Route-ID order was positive (ID+) whereas that in the ID-Route order was negative (R-). Because this order effect is a bias that distorts the normative values of conditional probabilities, it prevents the correct learning of the conditional probabilities. This order effect for actual decisions has not been reported previously. The Rescorla-Wagner model (Rescorla & Wagner, 1972; see also Gluck & Bower, 1988; Shanks, 1990b), one of the well-known models of frequency learning, predicts accurate acquisition of frequency information for multiple evidence items that are presented in parallel. It cannot account for the inaccurate acquisition of frequency information for multiple sequentially presented evidence items found in our present study. This is because the Rescorla-Wagner model completely ignores the temporal sequence of information. Thus, in order to account for the full spectrum of frequency learning, models of frequency learning such as the Rescorla-Wagner model should be modified to take into account temporal information. Hogarth & Einhorn’s (1992) anchoring and adjustment model, one of the well-known models of belief updating, can explain the recency order effect for actual decisions found in our present study with an extra assumption that subjects behaved in a step-by-step manner with implicit belief evaluations. However, the anchoring and adjustment model does not have any mechanisms for frequency learning. Thus, it cannot explain the learning behavior in our experimental tasks.

The second finding is a recency order effect for belief evaluations after frequency learning. This order effect for belief evaluations is a result directly predicted from Hogarth and Einhorn's anchoring and adjustment model. Many biases in decision making, such as the conjunction fallacy, overconfidence, representative heuristics, and so forth, would disappear with the correct learning of base rates from sequences of real events. However, the order effect for evaluations found in the present study, which is a special type of reasoning bias, cannot be eliminated even if base rates are used correctly. This result suggests that the mechanisms of base rate learning and use are relatively independent of the mechanisms of the order effect.

The third finding is the transfer effect across different base rate conditions. The base rate learned in one environment was transferred to another environment that had a different base rate. However, the transfer effect became weaker and weaker and soon diminished with the learning of the base rate of the new environment. Thus, with sufficient training, the base rate in a new environment could be eventually learned. This transfer effect of base rates is another demonstration of the finding that people are sensitive to base rates when frequencies of occurrence are presented in terms of sequences of real events. Although there was a transfer effect for base rates, belief evaluations carried out for one base rate environment were not transferred to a new base rate environment. This result is not surprising because belief evaluations for both environments were performed after the learning of base rates was complete. Once the learning was complete, the knowledge of the old environment was largely overridden by the knowledge of the new environment.

The experimental results of our present study can neither be solely explained by the existing models of frequency learning nor solely by the existing models of order effects because the former models do not consider sequential information whereas the latter models do not consider frequency learning. In order to account for the complete spectrum of frequency learning and order effects, we need a unified model that considers both frequency learning and order effects. We are currently working on such a model. Although we have not finished the model yet, we have some preliminary results, which are described in Wang, Johnson, & Zhang (1997). This model integrates the Rescorla-Wagner model for frequency learning and the Echo model for belief evaluations. Echo is the connectionist implementation of the Theory of Explanatory Coherence (Thagard, 1989, 1992), which is a theory of belief evaluations. In Echo, propositions (both hypotheses and evidence items) are represented by nodes, and coherent relations are represented by positive links and incoherent relations by negative links. The system runs in a parallel constraint satisfaction fashion. When the system settles down, the best hypothesis emerges as the one with the highest belief value. The original Echo ignores sequential information and does not have a learning mechanism, both of which are essential for a successful unified model of frequency learning and order effects. Thus, we added the Rescorla-Wagner rule into Echo so that the modified Echo has a learning mechanism. We also modified the original Echo so that it can accept sequential data for belief updating. With these two modifications, the modified Echo can learn frequency information from sequences of real events and can produce the order effect for belief evaluations. The simulation results from the modified Echo largely matched the empirical results from the experiments. However, there is still a phenomenon that the modified Echo cannot handle. This is the order effect for actual decisions, which is due to the sequential presentation of evidence items during learning. The learning mechanism in the modified Echo is based on the Rescorla-Wagner rule, which ignores temporal information. Therefore, the modified Echo cannot encode and learn the temporal information for sequentially presented evidence items. We are exploring alternative learning mechanisms that can learn frequency information both for evidence items that are presented in parallel and for evidence items that are presented sequentially.

REFERENCES

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica, 44,* 211-233.

Carroll, J. S. & Siegler, R. S. (1977). Strategies for the use of base-rate information. Organizational Behavior and Human Performance, 19, 392-402.

Cassecells, B. S., Schoenberger, A., & Graboys, T. B. (1978). Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine, 299,* 999-1000.

Christensen-Szalanski, J. J. J. & Beach, L. R. (1982). Experience and the base-rate fallacy. *Organizational Behavior and Human Performance, 29,* 270-278.

Christensen-Szalanski, J. J. J., & Bushyhead, J. B. (1981). Physicians' use of probabilistic information in a real clinical setting. *Journal of Experimental Psychology: Human Perception and Performance,* 7 (4), 928-935.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition, 58*, 1-73.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102*, 684-704.

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, *117* (3), 227-247.

Gluck, M. A., & Bower, G. H. (1990). Component of pattern information in adaptive networks. *Journal of Experimental Psychology: General, 119*, 105-109.

Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist, 39*(12), 1372-1388.

Hogarth, R. M. & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, 24, 1-55.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

Kleiter, G.D. (1994). Natural sampling: Rationality without base rates. In G.H. Fischer & D. Laming (Eds.), *Contributions to Mathematical Psychology, Psychometrics, and Methodology*. New York: Springer-Verlag.

Lyon, D., & Slovic, P. (1976). Dominance of accuracy information and neglect of base rates in probability estimation. *Acta Psychologica, 40*, 287-298.

Manis, M., Dovalina, I., Avis, N. E. and Cardoze, S. (1980). Base rates can affect individual predictions. *Journal of Personality and Social Psychology*, *38*, 287-298.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory*. New York: Appleton-Century-Crofts.

Shanks, D. R. (1990a). Connectionism and human learning: Critique of Gluck and Bower (1988). *Journal of Experimental Psychology: General, 119*, 101-104.

Shanks, D. R. (1990b). Connectionism and the learning of probabilistic concepts. *Quarterly Journal of Experimental Psychology*, *42A*, 209-237.

Thagard, P. (1989). Explanatory Coherence. *Behavioral and Brain Sciences*, 12, 435-502.

Thagard, P. (1992) *Conceptual Revolutions*. Princeton, New Jersey: Princeton University Press.

Towne, D. (1995). *CIC: Tactical Decision Making (Version 2.0)*. Behavioral Technology Laboratories, University of Southern California.

Wang, H., Johnson, T. R., & Zhang, J. (1997). *UEcho: A model of uncertainty management in human abductive reasoning*. Unpublished manuscript.