

## An encounter frequency account of how experience affects likelihood estimation

NATALIE A. OBRECHT, GRETCHEN B. CHAPMAN, AND ROCHEL GELMAN  
*Rutgers University, Piscataway, New Jersey*

When making judgments, people often favor information received from a few individual sources over large-sample statistical data. Individual information is usually acquired piece by piece, whereas statistical information combines many observations into a single summary. We examined whether this difference in the frequency of encounters affects how data are weighted. In two experiments, subjects read statistical information indicating an event to be rare and contrasting information from individual cases suggesting the event to be common. We controlled whether the individual cases were summarized into a single summary like statistical information, or presented serially, case by case. Subjects' estimates of event frequencies were higher when the individual cases were presented in serial, rather than summarized, format. A third study demonstrated that subjects treat each data sample as an instance, and do not weight according to sample size. These results support the conclusion that people weight information according to encounter frequency.

People are constantly faced with uncertainty. We can only estimate how likely we are to catch the flu this season, to experience medication side effects, or to miss the morning train. To judge how likely an event is to occur, information may be recalled from a variety of sources. In this article, we discuss such sources as belonging to one of two categories: statistical or individual. We aim to show that this distinction is useful when predicting how people will judge the likelihood of an event's occurrence.

We start with a familiar result—that is, people's tendency to overlook relevant statistical information. Fong, Krantz, and Nisbett (1986) demonstrated this phenomenon when they asked subjects to make judgments about a number of hypothetical problems in which statistical principles were relevant. For example, subjects read about the *Caldwells*, a couple trying to decide which of two car brands, Saab or Volvo, is better. Subjects read that both the *Consumer Reports* car experts and *Consumer Reports* readers slightly favored Volvos over Saabs. Then they read that the *Caldwells* have two Saab-owning friends and one Volvo-owning friend. The two Saab owners had few problems with their cars, whereas the Volvo owner describes negative experiences with his car. When asked to evaluate scenarios such as these, fewer than half of subjects' responses included a statistical explanation. Clearly, at least to psychologists, sample size is a relevant concept in this example. The information from *Consumer Reports* is based on many more people's experiences than that learned from the *Caldwells*' three friends.

Overlooking sample size when comparing information from statistical and personal sources is related to Nisbett and Ross's (1980) "man who," or, more recently, "person

who" (Stanovich, 2001), reasoning. "Person who" reasoning occurs when one discounts statistical information in light of a conflicting personal anecdote. For example, one may doubt a report stating that employment is on the rise, because he knows a "person who" can't find a job. This type of reasoning is problematic, because it runs counter to the law of large numbers. According to both the weak and strong versions of the law, a population mean is generally better approximated by a large, rather than a small, sample from a population. Sedlmeier and Gigerenzer (1997) argued that other mathematical theorems, rather than the law of large numbers, are more appropriate for justifying sample size use as normative. Nevertheless, the overall picture is the same: Sample size is a normatively relevant factor, but it is commonly overlooked. In the Volvo–Saab scenario, there is more reason to believe that the *Consumer Reports* data provide a reliable estimate of car soundness, because they are based on a larger sample than the information provided by the *Caldwells*' friends. From this viewpoint, the Fong et al. (1986) subjects should have explained that the Volvo is a better choice than the Saab because there is more evidence to support this hypothesis than the other way around, but they did not.

In a similar study, Nisbett, Krantz, Jepson, and Kunda (1983) asked subjects to read about David, who is considering attending either a liberal arts college or an Ivy League university. Subjects read that David has friends at each school and that his friends at the liberal arts college report more positive experiences at their school than do his friends at the Ivy League university. However, subjects then read that David visited each school for a day. At the liberal arts school, David is described as having a couple

---

N. A. Obrecht, natalie@ruccs.rutgers.edu

---

of negative experiences, whereas at the Ivy League school a few different positive experiences are described. After reading about the experiences of David and his friends, nearly three quarters of the subjects indicated that David should attend the Ivy League school. Thus, the information about David's single day of experiences was favored over the many experiences summarized by David's friends. Once again, subjects' choices were consistent with the small-sample personal information rather than with the choice favored by more data.

Ubel, Jepson, and Baron (2001) have also shown the strong effects of a couple of testimonials relative to statistical data. They provided subjects with a description of the heart condition angina, followed by two possible medical treatments. The first treatment, angioplasty, was described as a procedure only 50% effective but with a short recovery period. The second treatment, bypass surgery, was described as requiring much more recovery time than the angioplasty, but being 75% effective. After reading this information, some subjects were simply asked which treatment they would prefer if they had angina; 58% indicated that they would choose to have the bypass surgery. Another group of subjects received four testimonials in addition to the statistical information. Two of the testimonials described the experiences of people who chose the angioplasty treatment, one with a positive outcome in which the treatment worked, and the other with a negative outcome, in which the treatment was not effective. The other two testimonials were about two individuals who chose bypass surgery: One had a successful outcome, and one did not. When asked which treatment they would choose, only 30% of subjects receiving this additional information now indicated that they would choose the bypass surgery treatment. As with the previous experiments described, subjects' choices appear to have been greatly affected by a small number of anecdotes that countered large-sample statistical data. Together, these findings raise the question: Why do people appear to over-look valuable sample size information?

Ubel et al. (2001) suggested that information from individuals is different from statistical data in that it is more personal and affective. They suggested that this difference may make testimonials more salient than statistical data. When one relays the story of one's car breaking down, one usually provides a vivid account, which leads to stronger information encoding relative to dry statistical data referring to the same event (Fagerlin, Wang, & Ubel, 2005). This could explain why subjects give explanations (Fong et al., 1986) and make judgments (Nisbett et al., 1983) that appear ignorant of sample size considerations. Clearly, affect is an important factor that influences memory and judgment (Loewenstein, Weber, Hsee, & Welch, 2001).

However, another factor that distinguishes statistical from individual data is the frequency of encounters with exemplars from the kind in question. People get information from other people they talk to in sequence. This means they have numerous encounters for a given domain or topic. This is in stark contrast to statistical information, which sums up many instances of the domain at once, making for only one encounter, from the subject's viewpoint. This is a potentially relevant distinction, because we know that

humans track at least some types of event frequencies automatically (Hasher & Zacks, 1984). Hasher, Zacks, and colleagues have conducted numerous experiments showing that humans automatically track word frequency (Hasher & Zacks, 1984; Hasher, Zacks, Rose, & Sanft, 1987; Zacks, Hasher, & Sanft, 1982). This ability appears to be a core cognitive mechanism (Gelman, 1996), an assumption consistent with Hasher and Zacks's (1979, 1984) failure to find any individual differences between people of different ages and Zacks et al.'s failure to find an effect of intelligence.

If online frequency tracking extends broadly to many domains, it may explain why humans appear to place more weight on a few testimonials than on large-sample statistical summary data. Say, for example, that one tracks the frequency of instances in which she has heard information about Volvos. Suppose also that such information comes from both individual and statistical sources. If each encounter with the relevant information is encoded regardless of sample size, we can predict that information from individuals will be overrepresented relative to statistical data. In other words, if one talks to two Volvo owners who have had negative experiences with their cars, and also reads a report about thousands of Volvo owners who generally liked their cars, the person will have two instances in mind that are negative and only one that is positive. Of course, higher level cognition could override such a mechanism that ignores sample size, but we will suggest that the frequency count of raw encounters can play an important role in situations in which statistical data are compared with information from a few contrasting personal sources.

Here, we present three experiments in which we show that encounter frequency does indeed affect how likely people think events are to occur.

## EXPERIMENT 1

In Experiment 1, we manipulated encounter frequency by varying whether information about individuals was serial or simultaneous. If information about, say, four individuals were presented serially, it would result in four encounters, but if it were presented simultaneously, it should be encoded as only one encounter and consequently would have less impact on the final inference.

### Method

#### Subjects

For course credit, 200 undergraduates served as subjects.

#### Materials and Procedure

In small groups in a lab, subjects completed paper questionnaires. These included scenarios wherein a person was described as considering a choice with an uncertain outcome—for example, "Andrea will buy a stereo and wants to know how likely it is to break within a year because she could buy an extended warranty." After each description in the same spirit, subjects read a statistic indicating that the event of interest is rare (e.g., "of 1,000 stereos, 30, or 3%, broke within a year"). The statistics in the scenarios ranged from 1% to 6%. Subjects then read about individual cases, in which 75% of a small number of people experienced the event. These individual cases were presented either in a serial or a summarized format. The summarized format gave the outcome frequency in a single statement—for example, "Andrea talks to four stereo owners and three of their stereos broke

within a year.” The serial format stated each individual’s outcome one by one, but without providing any additional information—for example, “The first person said her stereo broke within a year. The second person . . .” and so on. To reduce any memory effects, these individual outcomes were given as separate lines on the same sheet of paper so that subjects could view all pieces of information at once. Subjects then indicated how likely the person in the scenario was to experience the event described (e.g., Andrea’s stereo breaking within a year) by responding with two dependent measures. First, they marked a point on a 9-point rating scale ranging from *extremely unlikely* to *extremely likely*; then they estimated the percentage chance (e.g., 30%) of the individual’s experiencing the event. (These two dependent measures gave highly similar results, so we discuss only the percentage chance measure in the Results section.) Additionally, subjects were asked to explain how they reached their conclusions. These responses were later coded for content favoring statistical or individual information, use of statistical or individual sample size, expressing uncertainty, and failure to interpret the scenario as intended. Finally, subjects were given a numeracy questionnaire similar to that used by Lipkus, Samsa, and Rimer (2001).

### Design

In addition to encounter format (serial, summary), we manipulated the number of individuals (0, 4, 16) within subjects to examine the potential role of sample size. Encounter format was crossed with the number of individuals in a  $2 \times 2 + 1$  design to make five conditions. The conditions were crossed with five scenarios (e.g., Andrea and her stereo) in a Latin-square design. Therefore, each subject read five scenarios, each providing a statistic indicating an event to be rare, followed either by no other information (i.e., 0-individual condition) or by information from 4 or 16 individuals, which was presented in either a serial or a summary fashion. We manipulated the presentation order of these scenarios on the basis of three dichotomous factors that we hypothesized could affect responses. This resulted in eight different orderings.

The first factor was whether the 0-individual condition, in which only statistical information was given, was seen before or after the four other conditions. We were concerned that subjects might pay extra attention to the statistical information in subsequent scenarios if this condition was seen first; therefore, half the subjects saw the 0-individual condition first and half saw it last. Next, we varied whether subjects saw the serial conditions before or after the summarized conditions. We were concerned that subjects would notice this change in presentation format and suppress any tendency they might have to give greater weighting to serially presented information. Finally, we manipulated whether subjects saw one of the 4-individual conditions before or after a 16-individual condition. Crossing the five conditions from the Latin square with the eight orders resulted in 40 versions of the questionnaire. We collected data from five groups of the 40 versions, resulting in 200 subjects total.

Our main prediction was that subjects would estimate events to be more likely when the individual information was presented serially than when it was summarized. We expected subjects’ explanations to correlate with their responses so that higher estimates would correspond to comments on the value of the individual information and the number of individuals providing counter cases. We predicted that numeracy would be negatively correlated with higher percentage estimates, which indicate a failure to account for the sample size of the statistical information.

### Results

The percentage estimate data were skewed and trimodal with modes around the statistical values given in the scenarios (e.g., 3%), 50%, and 75%. A power transformation of the data partially addressed these issues but yielded the same pattern of results as did analyses of the raw data. Thus, we simply report analyses of the untransformed data.

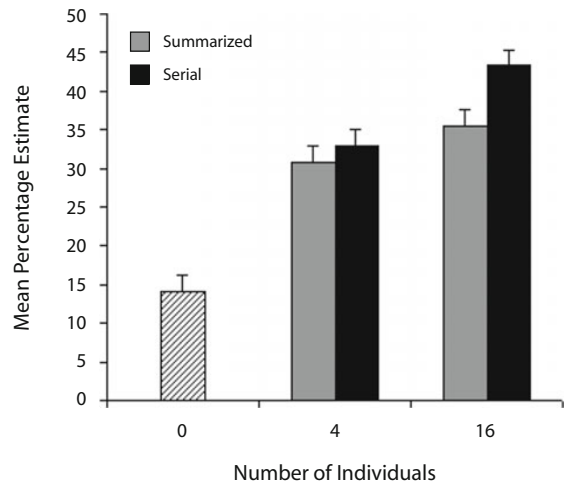


Figure 1. Main effects of encounter format and number of individuals on the percentage estimate measures from Experiment 1 (standard error bars are shown).

The pattern of reliable effects was consistent with our prediction that individuals’ use of data is influenced by whether it is presented in a summarized or a serial format.

### Percentage Estimates

We examined percentage estimate responses first in an ANOVA that excluded the 0-individual condition. This included the 2 (encounter format)  $\times$  2 (number) experimental manipulations, plus our Latin square (5 conditions) and 3 dichotomous counterbalancing factors described above. This analysis showed an effect of encounter format [ $F(1,157) = 9.64, p = .0023, \eta_p^2 = .0578$ ]. Thus, subjects’ estimates of the percentage chance of an event’s occurring were greater when the individual data indicating the event to be common were presented in a serial, rather than a summarized, fashion. Also, subjects judged the percentage chance of an event to be higher when they received information that conflicted with the low statistic from 16, rather than 4, individuals [ $F(1,157) = 25.71, p < .0001, \eta_p^2 = .1407$ ]. These main effects are shown in Figure 1. Additionally, presentation format and number interacted [ $F(1,157) = 4.40, p = .0376, \eta_p^2 = .0273$ ] in such a way that the effect of the serial presentation was greater when information from 16 rather than 4 individuals was provided. This interaction was qualified by a three-way interaction with version as well [ $F(4,157) = 2.86, p = .0252, \eta_p^2 = .0680$ ], indicating that the scenario in which the conditions were embedded affected the relationship between presentation format and number. These results are summarized in Table 1.

Next, we examined all conditions, including the 0-individual condition, and confirmed that subjects treated the five conditions differently [ $F(4,620) = 46.87, p < .0001, \eta_p^2 = .2322$ ]; events described by only low base rate statistical information were judged as less likely than events presented in scenarios involving individual counter cases [ $F(1,620) = 7.31, p < .0001, \text{preplanned contrast}$ ].<sup>1</sup>

**Table 1**  
Main Effects and Interactions of Interest  
From Experiment 1

Source	Percentage Estimate Data		
	<i>df</i>	<i>F</i>	<i>MS<sub>e</sub></i>
Between-Subjects Effects			
Latin-square version (V)	4,157	0.38	1,966.2
Order of 0 condition (O-zero)	1,157	2.40	1,966.2
Order of 4, 16 conditions (O-4/16)	1,157	0.80	1,966.2
Order of encounter format (O-EF)	1,157	3.54	1,966.2
Within-Subjects Effects			
Encounter format (EF)	1,157	9.64*	465.5
Number of individuals (N)	1,157	25.71**	405.1
EF × N	1,157	4.40*	440.9
EF × V	4,157	3.44*	465.5
N × V	4,157	4.33*	405.1
EF × N × V	4,157	2.86*	440.9

Note—The ANOVA included all main effects and interactions and excluded the 0-individual condition. Only the interactions of interest are shown here. These additional interactions were observed: V × O-zero × V × O-zero × O-4/16\*, O-zero × O-4/16 × O-EF\*, N × V × O-zero × O-4/16\*, N × V × O-zero × O-EF\*, N × O-4/16 × O-EF\*. \**p* < .05. \*\**p* < .0001.

### Explanations of Chance Estimates and Numeracy Data

Subjects' explanations of their percentage estimates were coded by a research assistant. There was 92% agreement between the assistant's coding and a subset of the responses coded by the first author. We used a mixed regression model to examine the extent to which subjects' explanations and numeracy levels predicted responses on percentage measure. Responses from the 0-individual condition were excluded, since some of the response codes were not applicable to this condition.

Subjects who judged events to be more likely showed a tendency to comment on the usefulness of the individual data ( $r = .36$ ). Surprisingly, subjects who gave higher estimates also showed a slight tendency to claim that the statistical data were useful ( $r = .18$ ). A qualitative look at the responses showed that some subjects would explain

**Table 2**  
Correlations Between Coded Explanations  
and Percentage Estimates

Factor	Percentage Estimate Data		
	<i>df</i>	<i>t</i>	<i>r</i> <sup>2</sup>
Favor stat	589	4.47**	.18
Favor indiv	589	9.26**	.36
Stat N	589	-18.88**	-.61
Indiv N	589	11.43**	.43
Uncertain	589	5.83**	.23
Fail	589	-1.58	.06

Note—A mixed model regression analysis yielded beta weights with corresponding *dfs* and *t* statistics. The  $r^2$ s reported here were derived using the formula  $r^2 = t^2 / (t^2 + df)$ . Favor stat, statistical information useful and/or better than the individual information. Favor indiv, individual information useful and/or better than the statistical information. Stat N, statistical sample size (e.g., 1,000 people) or probability (e.g., 2%) mentioned. Indiv N, individual sample size (e.g., 4 people) or probability (e.g., 75%) mentioned. Uncertain, mentions being uncertain about the chances of the event's occurring. Fail, fails to interpret the problem as intended. \*\**p* < .0001.

that they valued the statistical data but that they also felt they should consider the individual information. Such a comment was coded as favoring both types of information, which may have led to this relationship. Subjects who gave higher percentage estimates were not inclined to mention the statistical sample size ( $r = -.61$ ), but to point out the number of individuals described in the scenario ( $r = .43$ ). They also tended to express more uncertainty in their reasons ( $r = .23$ ) and to score lower on the numeracy measure ( $r = -.41$ ) than did subjects less biased by the individual counterexamples. The beta weights and *r* values in Table 2 depict the extent to which the six coded factors for subjects' reasons and subjects' numeracy levels predicted the percentage estimates that subjects assigned to the scenarios.

Additionally, we correlated subjects' sensitivity to encounter format with their numeracy. We quantified sensitivity to encounter format by subtracting each subject's average percentage estimate for the two serial scenarios from their average percentage estimate for the two summarized scenarios. These difference scores were not correlated with subjects' numeracy scores ( $r = .06, p = .3962$ ). Thus, subjects' bias to give greater weight to information presented piece by piece appears to be independent of numerical ability.

### Discussion

This first experiment demonstrated an effect of encounter format: Subjects' likelihood estimates were higher when the individual information was presented serially than when the same information was summarized. Thus, when there were many encounters indicating that an event was common, subjects rated the event as more likely than when there were few encounters contrasting with the low statistic. This result is especially interesting, because the difference between the encounter format conditions, serial and summarized, was minimal. When the outcomes of the 4 and 16 individuals were presented serially, they were written in such a way that each was stated on a single line on the same page. When these same data were summarized, they were presented together in a sentence. The difference between these two conditions is minuscule, compared with the analogous real-life situations into which we are trying to gain insight. One such example is when one receives information from talking to several different people serially, rather than reading about a study in a newspaper. However, even with our controlled small-scale manipulation, we demonstrated an effect of encounter frequency.

In addition to this main finding, we found that subjects were so sensitive to the number of counter cases that they judged events to be much more likely when there was any such information from individuals. This large leap, as seen in Figure 1, is shown by the difference in subjects' estimates when there were 0 rather than 4 or 16 individuals providing information, in addition to the statistic. Also, subjects were so sensitive to the number of counter cases that they judged an event to be more likely to occur if 12 out of 16, rather than 3 out of 4, people reported experiencing the event of interest. Given that the statistic in each scenario summa-



alized at least 1,000 people's experiences, these increases in likelihood judgments are extreme. If only 3% of 10,000 people experience Event A, additional information about the outcomes of 4 or 16 people should matter little.

Also, encounter frequency and number of individuals interacted, so that the effect of the serial presentation was greater when there were 16 rather than 4 individuals. This interaction is consistent with our account that subjects are tracking the number of encounters they have with information relevant to an event.<sup>2</sup>

Subjects were asked to explain how they decided the chances of each event's occurrence. Of the six coding factors that we employed, three accounted for the most variance. Subjects who commented on the value of the individual data, and mentioned the number of such counter cases, tended to give higher estimates, whereas those who mentioned the statistical sample size tended to give lower ones.

Finally, we found a relationship between subjects' numerical abilities and their evaluation of how likely an event was to occur. More numerate subjects gave lower estimates, indicating that they were less influenced than were less numerate subjects by the small number of counter cases provided by the individual information. It is interesting to note that we failed to find a relationship between numeracy and sensitivity to encounter format; that is, numerate subjects' event likelihood estimates were as biased by the serial presentation as were the estimates of those lower in numeracy.

In summary, these results indicate that serial presentation causes an increase in the weight given to information. We suggest that this is because encounter frequency is tracked by a low-level cognitive system, so therefore serial presentation leads to increases in the number of instances acquired in a category. One possible challenge to our finding is that a recency effect is responsible for this result. If subjects' responses were primarily based on the last few pieces of information that they had read, we would predict that they would be more influenced by the individual information when it is spread out in the serial condition than when it is summarized. This is because in the summarized case subjects read statistical information followed by a single summary statement, whereas in the serial condition the statistical information was followed by 4 or 16 statements. To see whether the encounter format effect could be explained in this way, we conducted a follow-up experiment that manipulated the order of the two types of information.

## EXPERIMENT 2

In our second experiment, we addressed the concern that our initial finding was due to a recency effect by manipulating the presentation order of the statistical and individual information. Also, we were interested in examining the emotional salience of the individual information to see how this might affect subjects' weighting of information. We expected subjects to judge events to be more likely when they read salient testimonials from individuals describing their experience than when the same outcome was simply stated, as in Experiment 1.

## Method

### Subjects

For course credit, 417 undergraduates served as subjects.

### Materials and Procedure

This was an Internet-based experiment. We replicated the first experiment by manipulating encounter format (serial, summary) and number of individuals (4, 16) within subjects. These factors were crossed to make four conditions, which were then crossed in a Latin square with four scenarios. (The 0-individual condition was not used in this experiment.) The presentation order of the four scenarios was counterbalanced across subjects. Again, we asked subjects to indicate how likely they thought the event of interest was to occur by marking a point on a 9-point rating scale and by providing a percentage estimate (e.g., 10%). Again, these two measures gave nearly identical results; we report only the percentage data.

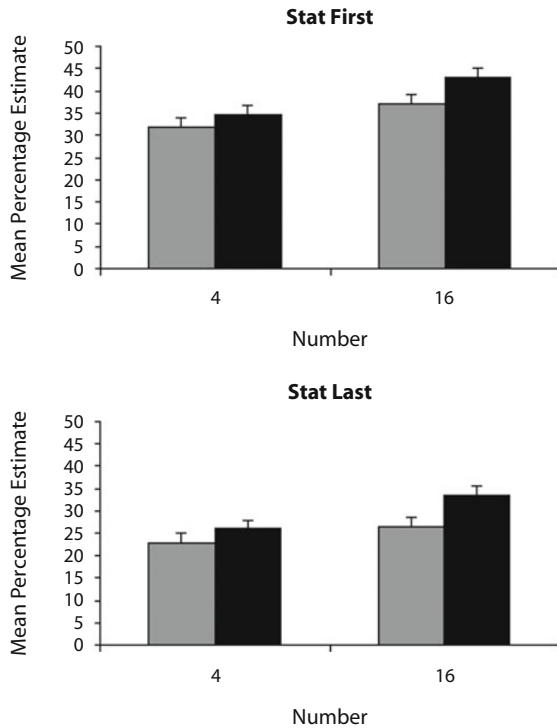
In addition to these original factors, two new variables were manipulated. The presentation order of the statistical and individual information within a scenario was varied between subjects. Thus, half of the subjects always read the statistical information at the start of each scenario, followed by the individual information, as in Experiment 1, whereas the other half always received the individual information before the statistical. Each piece of information was presented on a different Web page, such that the serial condition with 4 individuals consisted of 5 pages, 1 giving statistical information and 4 pages for the 4 individuals. This design should have boosted any recency effect, relative to the first experiment, for subjects in the statistics-first condition because in the serially presented scenarios the statistical information was presented 5 or 17 Web pages prior to the response page with the dependent measures. In contrast, when subjects were reading the summarized scenarios, the statistical information was only 2 Web pages back, because the information from individuals was summarized on a single page. If the difference between the serial and summary conditions in Experiment 1 can be explained by a recency effect, we should not expect to replicate the encounter frequency finding with subjects in the individual-information-first condition.

Next, we manipulated the emotionality of the serially presented individual information, such that one group of subjects received individual information with no emotional content, as was the case in our first experiment, whereas the other group received more descriptive testimonials describing how the event of interest in the scenario did or did not occur. For example, in the scenario in which Andrea talks to stereo owners because she is trying to learn how likely a stereo is to break, testimonials like the following were used:

The first person reported, "The Sony stereo I bought for my niece stopped working not long after I gave it to her. I felt bad that her gift didn't last longer." The next person said, "I've been very happy with my Sony stereo. I use it every morning while I'm getting ready for school. It is still working great and I've had it for a number of years." The third person explained, "My Sony stereo gave me trouble not long after I bought it. I was using it to listen to lectures I recorded from my classes and it stopped working right when I was reviewing for an exam."

This manipulation allowed us to experimentally examine the assumption made by other researchers: that the emotional salience of individual testimonial information plays a role in this type of information's being given greater weight than statistical data.

Finally, we collected individual differences data. As in Experiment 1, we asked subjects to complete a short numeracy questionnaire which included converting among probabilities, proportions, and percentages. In addition to this, we used Frederick's (2005) cognitive reflection test (CRT), and asked subjects their gender, SAT quantitative (SAT-Q) score, and whether they had ever taken a statistics class or had had lessons in probability theory.



**Figure 2.** Effects of encounter format and number of individuals on the percentage estimate measure from Experiment 2 (standard error bars are shown). The top graph shows when the statistical information was given first, and the bottom graph shows when it was presented last.

## Results

### Percentage Estimates

As in Experiment 1, the pattern of results was consistent with our encounter frequency account [ $F(1,389) = 17.47$ ,  $p < .0001$ ,  $\eta_p^2 = .0430$ ]. As shown in Figure 2, subjects were more influenced by information from individuals that

contradicted statistical data when the individual information was presented serially rather than in a summarized format (see Table 3). Also, they gave higher percentage estimates when there were 16, rather than 4, counter cases [ $F(1,389) = 31.33$ ,  $p < .0001$ ,  $\eta_p^2 = .0745$ ]. Both encounter format and number of individuals interacted with version [ $F(3,389) = 3.36$ ,  $p = .0189$ ,  $\eta_p^2 = .0252$ ;  $F(3,389) = 6.54$ ,  $p = .0003$ ,  $\eta_p^2 = .0480$ ], indicating that the strength of these main effects partly depended on the scenarios that they were imbedded in. Also, an interaction between presentation format, number of individuals, and version was uncovered [ $F(3,389) = 10.12$ ,  $p < .0001$ ,  $\eta_p^2 = .0724$ ]. An examination of the data shows a trend toward a larger effect of encounter frequency when there are 16 individuals, but this relationship is influenced by version.

An effect of the presentation order of the statistical information [ $F(1,389) = 21.12$ ,  $p < .0001$ ,  $\eta_p^2 = .0515$ ] indicates that events were rated as less likely when the low statistical base rate was viewed just before the likelihood judgment; that is, we found a recency effect. Importantly, however, we did not find an overall two-way interaction between encounter format and order of the statistical information, indicating that the encounter format effect was no smaller when the low statistical base rate was viewed just before the likelihood judgment. This finding was followed up by an ANOVA that examined only data from subjects who received the statistical information just before judging the likelihood of an event's occurrence. We still found an effect of encounter format [ $F(1,196) = 8.10$ ,  $p = .0049$ ,  $\eta_p^2 = .0397$ ], indicating that the effect of serially presented information cannot be accounted for by the recency effect. Graphs of the main effects for each presentation order are shown in Figure 2.

In the general analysis, we failed to find an effect of description level of the testimonials [ $F(1,389) = 0.66$ ,  $p = .4177$ ]; this factor did not interact with encounter format, either [ $F(1,389) = 1.07$ ,  $p = .3015$ ]. This was surprising, because we had expected subjects to rate events as more likely when individuals provided detailed accounts of how statistically rare events happened in their lives. To narrow

**Table 3**  
Main Effects and Interactions of Interest From Experiment 2

Source	Percentage Estimate Data		
	<i>df</i>	<i>F</i>	<i>MS<sub>e</sub></i>
Between-Subjects Effects			
Presentation order of statistic (O-stat)	1,389	21.12**	1,650.13
Description level (DL)	1,389	0.66	1,650.13
Latin-square version (V)	3,389	1.07	1,650.13
Within-Subjects Effects			
Encounter format (EF)	1,389	17.47**	543.05
Number of individuals (N)	1,389	31.33**	495.37
EF × N	1,389	2.90	544.56
EF × O-stat	1,389	0.14	543.05
EF × V	3,389	3.36*	543.05
N × V	3,389	6.54*	495.37
EF × N × V	4,389	10.12**	544.56
DL × EF	1,389	1.07	543.05

Note—The ANOVA included all main effects and interactions. Only the interactions of interest are shown here. The only significant higher order interaction was O-stat × DL\*. \* $p < .05$ . \*\* $p < .0001$ .

in on the description level factor, we conducted a more sensitive ANOVA that focused only on responses to the descriptive and nondescriptive accounts (only the serial conditions). Again, we failed to find evidence of an effect of emotional testimonials [ $F(1,394) = 0.01, p = .9082$ ]. Subjects did not appear to give the individual counter cases more weight when they contained emotional wording and descriptive content than when they were dry and without detail.

### Individual Difference Measures

We conducted three sets of analyses using the individual difference measures (see Table 4). First, we correlated subjects' average percentage estimates across the four scenarios to their numeracy, CRT score (Frederick, 2005), SAT-Q score, gender, and exposure to statistics and/or probability theory. As in Experiment 1, less numerate subjects tended to estimate events to be overall more likely than did more numerate subjects (see Table 4), although this relationship was marginally significant using our corrected alpha ( $p = .0058$ ). Additionally, subjects with lower CRT scores gave higher percentage estimates.

Next, we correlated our six individual difference measures with subjects' encounter frequency bias. This was measured, as in Experiment 1, by subtracting each subject's mean percentage estimates for the serially presented scenarios from their mean percentage estimates for the summarized scenarios. Higher values indicate greater sensitivity to encounter frequency. Again, as in Experiment 1, we found no relationship between subjects' numeracy levels and their sensitivity to encounter frequency. Also, we failed to find a relationship between encounter frequency sensitivity and any of our five other individual difference measures.

Finally, we examined the correlations among our six individual difference measures. Subjects higher in numeracy tended to be male, score higher on the CRT, report higher SAT-Q scores, and have more exposure to statistics. Subjects who performed well on the SAT-Q also reported a background in statistics and probability theory. Lastly, those with statistics exposure generally had been exposed to probability theory.

### Discussion

Our second experiment showed that the effects of encounter frequency cannot be accounted for by a recency effect. Subjects did judge events to be more likely when they viewed the individual, not the statistical, information just prior to the dependent measures. However, this downward shift to overall lower estimates did not affect the magnitude of the difference between the serial and summarized conditions, as shown in Figure 2. When considering only the cases in which the statistic was viewed prior to the individual information, we still found an effect of encounter frequency. Therefore, information presented serially is given more weight than when it is summarized, regardless of the order in which it is received.

The effect of the number of individuals found in Experiment 1 was also replicated. Again, subjects judged events to be more likely when they viewed information from 16, rather than 4, people, 75% of whom reported experiencing the event.

Interestingly, we failed to find an effect of the description level of the individual information. Subjects judged events to be equally likely, whether they viewed descriptive testimonials from individuals or short, dry statements from each person. It is unlikely that we lacked the statistical power to detect this effect, given our large sample size, but our testimonials might not have been individuating or emotional enough. In order to keep the description level manipulation controlled, we did not give names to the individuals or show pictures of them. Perhaps if we had done so we would have found the predicted effect; but it is interesting that the effect of testimonials does not appear to be as robust as expected.

As in the first experiment, we found that subjects with higher numerical abilities tended to give overall lower percentage estimates than did those who were less numerate. Additionally, those giving lower percentage estimates tended to score higher on the CRT, so it appears that more mathematically savvy and reflective thinkers are less influenced by individual cases that counter statistical information.

Also consistent with Experiment 1, we failed to find a relationship between encounter frequency and numeracy. Numerate subjects were no less susceptible to the tendency to give greater weight to serially presented information,

**Table 4**  
Relationship Among Percentage Estimates  
and Individual Difference Measures From Experiment 2

	Mean Percentage	Encounter Frequency Bias	CRT	SAT-Q	Gender	Statistics Exposure	Probability Exposure
Numeracy	-.15	-.02	.39*	.42*	-.18*	.19*	.13
CRT	-.21*	.05	-	.45*	-.29*	.18*	.12
SAT-Q	-.11	-.10	-	-	-.15	.23*	.17*
Gender	.06	.02	-	-	-	-.15	-.12
Statistics exposure	-.12	-.10	-	-	-	-	.34*
Probability exposure	-.03	-.00	-	-	-	-	-

Note—Mean percentage refers to subjects' judgments regarding the events' occurring across the four scenarios. Encounter frequency bias was found by subtracting the mean percentages each subject gave in the summarized conditions from their judgments in the serial conditions. CRT, cognitive reflection task; SAT-Q, SAT quantitative scores. \* $p < .0013$  (39 correlations total;  $\alpha = .05/27 = .0019$ ).

relative to summarized information, than were less numerate subjects. We also failed to find any relationships between an encounter frequency bias and scores on the CRT or SAT-Q, gender, or exposure to statistics or probability. Again, this fits with the idea that the encounter frequency effect arises due to a low-level cognitive mechanism not greatly affected by individual differences.

### EXPERIMENT 3

In Experiments 1 and 2, we controlled whether data were presented case by case or in a single instance to show that sequential data are weighted more than are data summarized together. It is not clear whether the encounter frequency effect is specific to the influence of individual cases. If people base their likelihood estimates on the number of encounters with no attention to sample size, we should find analogous effects, even if none of the information presented to subjects is an individual case or testimonial. We tested this prediction in Experiment 3. Specifically, we held constant the number of encounters, but varied how different outcomes were paired with different sample sizes. Unlike our first two experiments, each encounter consisted of a group of outcomes. As a consequence, we used probabilistic data, rather than the dichotomous outcomes. The prediction afforded by our encounter frequency account was that encounters would be weighted equally, regardless of sample size. This prediction is in contrast to a more normative model that states that people weight data according to sample size.

#### Method

##### Subjects

For course credit, 99 undergraduates participated.

##### Materials and Procedure

This computer-based study was run in a lab. Subjects were asked to imagine that they were nature preserve workers trying to learn about different types of animals, and that at a zoology conference, they talked to other nature preserve workers about their experiences with various animals—for example, “There are hyenas at the nature preserve. You notice that some hyenas have long fur, but others have short fur. You talk to other nature preserve caretakers about how often they’ve seen long-haired hyenas.” Subjects then read percentage and sample size information from six different individuals regarding the animal of interest (e.g., “One of the caretakers tells you that of the 750 hyenas he has seen, 20% had long hair. Another nature preserve worker says . . .”). As in Experiment 2, each statement was given on a separate Web page. Subjects were then asked to estimate the chances that the next animal born on their nature preserve of the given type would have the characteristic of interest (e.g., that the next hyena born on the nature preserve would have long hair). They used a 9-point rating scale and gave a percentage estimate for each of six animals. As in our previous experiments, these dependent measures gave nearly identical results, so only the percentage measure is reported. We chose animals that subjects would likely be familiar with, but that they would be unlikely to know much about.

Finally, we gave subjects a numeracy scale (Lipkus et al., 2001) similar to that used in Experiments 1 and 2, and collected information on exposure to probability theory and statistics, as well as gender.

##### Design

The experimental stimuli consisted of six data sets that were combined with six different animal stories. Each data set consisted of six probabilities, each with a corresponding sample size. For

example, one of the data sets gave the probability of an event to be 20% based on a sample of 750, 15% based on a sample of 1,000, 30% based on 100 cases, and three other percentage–sample-size combinations. According to the encounter frequency account, subjects should weight these probabilities equally (that is, average together 20%, 15%, 30%, etc.), whereas a more normative account would predict that they would weight the data according to sample size.

Across the six data sets, we manipulated, within subjects, the percentages in the data sets to be low, medium, or high in value. Two of the six data sets gave percentages in a low probability range, from 0% to 40%; the example above was from this range. Two data sets gave percentage data in a middle range, from 30% to 80%, whereas the remaining two provided data in a high probability range, from 70% to 100%.

Our main variable of interest was percentage–sample-size pairing. For three of the six data sets, one from each probability range, there were more instances toward the high end of the range showing the event of interest to be likely, but these were paired with smaller sample sizes. In the remaining three data sets, there were more instances suggesting that the event likelihood was toward the low end of the range, but these values were paired with smaller sample sizes. This method allowed us to contrast the encounter frequency and normative models, because they gave opposite predictions regarding the two types of data sets within each of the probability ranges (see Figure 3). The encounter frequency model predicts that subjects will give lower percentage estimates when they see more low than high percentages, whereas a normative model predicts that they will give higher estimates in this case, because higher percentage cases are paired with larger samples. If subjects considered both the frequency of cases and sample size, our manipulation of percentage and sample size pairing would have no effect, and thus might not show support for either model.

Both the data range and percentage–sample-size pairing were varied within subjects. Thus, all subjects viewed each of the six data sets. The pairing of data set and story (e.g., hyenas with short vs. long fur) was varied between subjects in one of two pairings. Also, the order of the stories (Order 1 or 2) and the presentation of the data within each story (Order 1 or 2) were varied between subjects, so we had eight between-subjects conditions in total.

#### Results

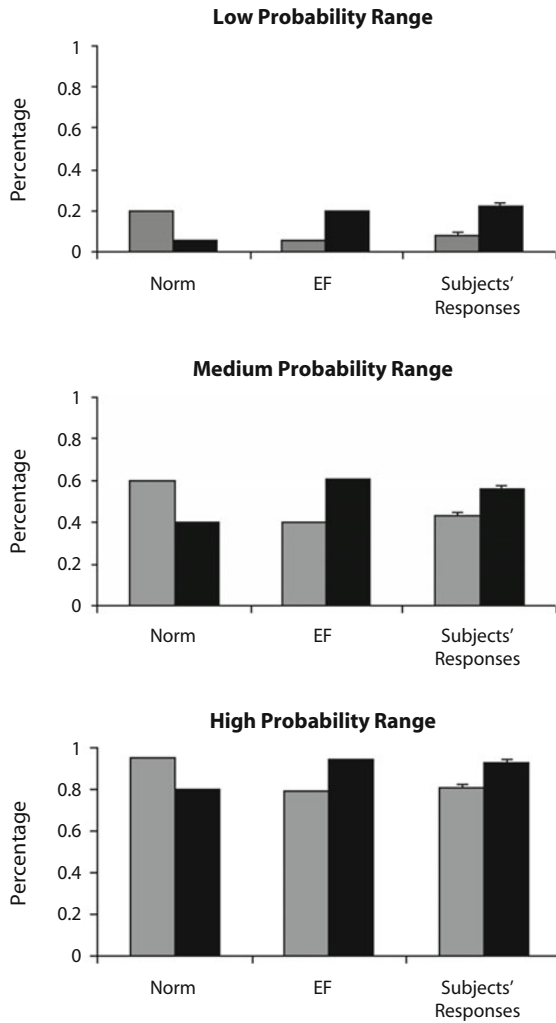
##### Percentage Estimates

We found an effect of percentage–sample-size pairing, as predicted by the encounter frequency account [ $F(1,90) = 127.72, p < .0001, \eta_p^2 = .5866$ ]. Subjects’ percentage estimates were higher when a data set provided more large-percentage instances, despite these instances being paired with smaller sample sizes. Also, subjects were highly sensitive to the probability range of the data [ $F(2,180) = 2,836.35, p < .0001, \eta_p^2 = .9692$ ], such that their estimates closely matched whether the data were in the low-, medium-, or high-range categories. Both of these main effects are apparent in Figure 3. Additionally, there were a few interactions involving our counterbalancing factors that are not of major theoretical interest.<sup>3</sup>

##### Measures of Individual Differences

Again, three of the stories given to subjects paired higher percentages with lower sample sizes, whereas the other three paired lower percentages with lower sample sizes. For each subject we calculated the average difference between their responses for these two levels of the percentage–sample-size pairing factor. If subjects gave equal weight to all percentages given in each of the six





**Figure 3.** Normative and encounter frequency model predictions and actual percentage estimate data of the six data sets used in Experiment 3 (standard error bars are shown). The gray bars show the model predictions and percentage estimates when data sets gave more low percentages, but paired them with low sample sizes. The black bars show data sets when more high percentages were given but were paired with low sample sizes.

stories, they would estimate the percentage chance of the events in the first set of three stories described to be, on average, 16% higher than the latter set of three stories. If subjects weighted the percentages according to sample size, they would instead weight the first set of stories as being 16% less likely than the second set. Using a median split, we found that low-numerate subjects almost perfectly fit with the encounter frequency prediction: They weighted events described by higher percentages paired with lower sample sizes to be 17.04% ( $\eta_p^2 = .8298$ ) more likely than events described by low percentages paired with smaller samples. High-numerate subjects also showed this pattern, but instead they judged the first type of events to be only 9.14% ( $\eta_p^2 = .3689$ )

more likely than the latter. Thus, less numerate subjects had a greater tendency to ignore sample size information than did subjects scoring higher on the numeracy scale ( $r = -.3439, p = .0005$ ). Table 5 summarizes the relationships among the individual difference measures and the extent to which subjects neglected sample size.

**Discussion**

Experiment 3 tested whether subjects' probability estimates are consistent with a normative account that weights data according to sample size or an encounter frequency account that gives data instances equal weight. Subjects' percentage estimates were highly similar to that predicted by the latter (see Figure 3). This experiment adds to the findings of the first two experiments by showing that, when integrating data, people give each piece of data very similar weight, despite differences in the sample size of those data. Also, this study shows that the encounter frequency idea applies to probabilistic data. Subjects considered percentage data rather than the individual dichotomous outcomes used in Experiments 1 and 2.

However, unlike our first two experiments, Experiment 3 showed that numeracy relates to how subjects integrate probabilistic information. More numerate subjects gave slight consideration to sample size, unlike less numerate subjects, who appeared to ignore it completely. Perhaps we found this effect here and not in Experiments 1 and 2 because in this experiment sample size information was explicitly mentioned, separate from the outcome data. For example, in the earlier experiments, subjects read about individual outcomes presented one after another or about three out of four people, in contrast to 15% of 1,000 cases. The sample size of four is necessary to interpret the three cases as constituting 75% of the total, whereas the value 15% can stand alone. This may prompt more numerate subjects to infer that perhaps sample size is a relevant piece of information (Grice, 1975).

Overall, this experiment provides support for our encounter frequency account. When deciding how likely an event is, it appears that people rely on the average of their experiences, to the complete or near exclusion of sample size information.

**Table 5**  
Relationship Among Percentage Estimates and Individual Difference Measures From Experiment 3

	Sample Size Neglect	Gender	Statistics Exposure	Probability Exposure
Numeracy	-.34*	-.21	.16	.27
Gender	.26	-	-.24	-.08
Statistics exposure	-.24	-	-	.30*
Probability exposure	-.17	-	-	-

Note—Sample size neglect refers to the mean difference between subjects' judgments of the two levels of percentage-sample size pairing; on one level, more large percentages were paired with smaller samples, and on the other, more small percentages were paired with smaller samples. Correlations in the .20–.29 range were marginally significant. \* $p < .005$  (10 correlations total;  $\alpha = .05/10 = .005$ ).

## GENERAL DISCUSSION

The data we have presented here are consistent with our theory of encounter frequency. In the first two experiments, we showed that subjects give more weight to information presented piece by piece than to the same data given in a summarized statement. Specifically, subjects' judgments about the likelihood of an event were shifted further away from a low statistical value when they read about people who experienced the event individually, rather than when people's experiences were summarized together. We found this effect to be independent of the order of individual versus statistical information, and it does not appear to be affected by the descriptiveness of the individual data.

In Experiment 3, we tested whether people weight data equally according to the encounters they experience or according to sample size. We found that subjects' estimates of event likelihoods are consistent with an encounter frequency account; that is, they give instances nearly equal weight, despite data being based on vastly differing sample sizes. Although we assume that other factors than encounter format influence the weight that data are given, the experiments presented here show that encounter format itself does affect how people weight data. This finding contributes to an explanation of why people so commonly overlook sample size, even to the extent of undervaluing statistical data in the face of a few contradictory anecdotes.

Although our findings show a general effect of encounter frequency, we found mixed results across Experiments 1 and 2 regarding our prediction that this effect would be greater when there were more cases that countered the low statistical value. Because numerous studies support the position that humans have a nonverbal numerical representation system (e.g., Dehaene, Dehaene-Lambertz, & Cohen, 1998; Hurewitz, Gelman, & Schnitzer, 2006) and that people automatically track at least some frequencies (e.g., Hasher & Zacks, 1984), we suspected that people would be sensitive to the number of encounters they have had with a given category. Such a system would be very useful, but could lead to sample size neglect in an environment composed of encounters that differ in how much information they provide (e.g., statistical summary data vs. single-case examples). We expected the conditions with 16 individuals with their data presented serially to yield a greater encounter frequency effect than the 4-individual serial condition because, in the 16-individual condition, subjects should have had 5 encounters indicating that the event was rare (the statistic + 4 individuals) and 12 encounters indicating that the event was common (12 individuals reported experiencing the event). In contrast, in the 4-individual serial condition, subjects should have had 2 encounters indicating that the event was rare (the statistic + 1 individual) and 3 encounters indicating that the event was common (3 individuals reported the event). If subjects encode the frequency of encounters as we have suggested, they may compute the difference or ratio in the number of encounters they have had showing that an event occurs often or rarely. Either the difference or the ratio would be more extreme in the 16-individual condition (5 vs. 12) than in the 4-individual condition (2 vs. 3). In Experiment 1, we

found the predicted interaction between encounter format and number. In Experiment 2, this interaction fell short of significance ( $p = .0895$ ), so the issue of representation remains an interesting future question.

In addition to our main effect, we found other effects of interest. To start with, in Experiment 1, subjects' percentage estimates were surprisingly high in the condition in which they were given only statistical information. We had expected that subjects would simply read the scenario that included the base rate of an event, and, when asked the percentage chance of the event's occurrence, would simply copy down the percentage they had just read. However, a sizable portion did not do this. The statistical base rates given to subjects ranged from 1% to 6% in the five scenarios used, but the mean percentage estimate in this statistic-only condition was more than 14%. Thus, although many subjects estimated the chances of an event to be equal to that statistical base rate, a surprising number gave higher estimates. Looking at subjects' explanations, it was not uncommon to see statements like "It's unlikely to happen, so I guessed 15%." If subjects' explanations are taken at face value, it seems that some of them formed an approximate representation of the statistic (e.g., low chance) and reinterpreted this gist as a higher value (e.g., 15%). Since the statistics given in the first two experiments were low values, it makes sense that reinterpreted gists would be, on average, greater than the actual statistics, because the lower end of the percentage scale is bounded by 0. Reyna and Brainerd's (1995) fuzzy trace theory purports that people prefer to reason using gist representations, including when they are considering numbers. This is consistent with our account that subjects are using an alternative representation, such as the frequency of encounters, rather than the exact numbers given, to estimate the chances of an event's occurring.

It was also interesting that subjects showed extreme neglect of sample size. In Experiment 3, less numerate subjects completely ignored sample size, whereas more numerate subjects showed only a slight consideration of this factor. In Experiments 1 and 2, subjects greatly increased their event likelihood judgments when only a few individual cases countering a low statistic were given, whether these cases were given in a summarized or sequential format. Percentage estimates of how likely an event was to occur doubled when just a few individuals' information was given, in addition to statistical data that summarized anywhere from 1,000 to 100,000 data points. Another jump in likelihood estimation was seen in comparing responses to the conditions in which subjects read about 16, rather than 4, individuals. Since all of the statistics were based on very large samples, this sensitivity to a few counter cases seems drastic. This is especially surprising, because we purposely kept the individual information impersonal so that it provided no additional information, as compared with the statistical data. Andrea did not determine how likely her new stereo was to break by asking her friends or family (who might have idiosyncratic characteristics similar to hers); she talked to strangers who happened to be in the electronics store with her. Subjects may use a heuristic approach in which information from a few local strangers is more trustworthy than statistical

data that might represent only people residing in a distant place, and they might apply this heuristic approach even when there is no reason to believe that the individual cases are more informative than the statistical data are. In any event, this increase in likelihood estimation is in line with our encounter frequency account, which predicts that summarized statistical data will be treated in a manner similar to a single datum, such that a few cases can greatly shift one's perception of the chances of an event.

Next, we found that subjects higher in numeracy and reflective thinking were less biased by anecdotes that contradicted a low statistical base rate; that is, when compared with subjects lower in numeracy and reflective thinking, these subjects gave overall lower event likelihood estimates that were closer in line with the large-sample statistical data. This is not surprising, given the growing literature relating numeracy and judgment (Obrecht, Chapman, & Gelman, 2007; Peters et al., 2006), and also given that higher cognitive ability has been associated with use of sample size information (Kokis, Macpherson, Toplak, West, & Stanovich, 2002).

However, when we correlated our individual difference measures with subjects' degree of encounter frequency bias, we found no relationships in Experiments 1 and 2. In Experiment 3, we found that more numerate, statistically trained male subjects tended to pay slight attention to sample size. This fits with our expectation that persons trained in research methodology and/or statistics, such as psychology graduate students, would be more likely to override this encounter frequency effect with their inclination to consider sample size (Lehman, Lempert, & Nisbett, 1988). In Experiment 3, but not in the first two experiments, sample size information was explicitly given as a number separate from the outcome information (e.g., "out of 1,000 hyenas, 10% . . ." vs. "3 out of 4," or "The first person reported . . ."). This difference may have cued more numerate subjects to give some consideration to sample size, unlike less numerate subjects, who completely ignored it, as predicted by our account. It is possible that the right content domain could trigger even innumerate subjects to show some sensitivity to sample size (Nisbett et al., 1983; Obrecht, Chapman, & Suarez, 2009).

Our encounter frequency account relates to work that distinguishes between receiving risk information from description versus experience. Gottlieb, Weiss, and Chapman (2007) and Hertwig, Barron, Weber, and Erev (2004) have shown that subjects make different decisions depending on whether they learn about probabilities by simply being told what they are (e.g., 2%) rather than by sampling individual events to learn from experience. Thus, the format in which information is given is an important consideration when studying decision making and statistical reasoning alike.

Similarly, Brase and Barbey (2006), Cosmides and Tooby (1996), and Gigerenzer and Hoffrage (1995) claimed that humans should be sensitive to natural frequencies, since this would have been the primary method of obtaining probabilistic information during our evolution. Our belief is that this evolutionary explanation may apply better to experienced information than to summarized natural frequencies. This is because, in our first two experiments, subjects' judg-

ments of event frequencies were shifted further from the low statistic when they saw counter cases presented piece by piece, as they would be in the natural world, than when those same data were summarized using natural frequencies (i.e., 3 out of 4; 12 out of 16). This suggests that data acquired via natural sampling are not treated the same as identical data summarized in a natural frequency format. In our experiments, the former were given greater consideration than the latter.

Our encounter frequency account also has parallels with the validity effect (Hasher, Goldstein, & Toppino, 1977), which shows that repeatedly presented statements are judged to be more valid than statements presented only once. Our sequential presentation format did not repeat information over and over again, which could prompt subjects to infer that it was especially important; it gave different pieces of information one by one, rather than providing all of the data in a single instance. Thus, we did not tell the subject four times that a given person's stereo had broken, but instead relayed the experiences of four separate individuals. Additionally, our results in Experiment 3 are predicted by the encounter frequency account but are beyond the scope of the validity effect, since we held the number of presentations constant. Nevertheless, both our account and the validity effect show that people are sensitive to the frequency of encounters they have with information. It is possible that the encounter frequency effect is the mechanism underlying the validity effect, if subjects encode repeated presentations of the same information as separate encounters.

In this article, we have argued that when laypeople acquire information about the chances of an event, they tend to encode each experience, even if that experience is composed of many instances, as if it provided a fixed amount of information. In other words, we think that people are paying little, if any, attention to sample size; instead, they just keep track of the outcomes they experience. In an environment without statistics, this simplification of information makes sense; there is no need to consider the sample size of an outcome if you are dealing with single events. We suspect that the encounter frequency effect is automatic, but a definitive test awaits further research. Also, we point out that it must be the case that the tendency to weight data by encounters can be modified or overridden by higher level cognition, since this effect was shown to be less extreme in more numerate subjects in Experiment 3.

People's tendency to overlook sample size when comparing information from statistical and individual sources has been shown time and again (Borgida & Nisbett, 1977; Fong et al., 1986; Nisbett et al., 1983; Nisbett & Ross, 1980; Stanovich, 2001). The overweighting of individual data has been explained by its supposed salience and emotionality, as compared with dry statistical information. However, statistics often summarize many data in a single encounter, whereas individual experiences provide relatively little information but may total many encounters. We suggest that this imbalance in the number of encounters relative to the amount of data prompts judgments that favor individual instances over statistical data. This perspective sheds some light on situations in which one appears to overlook sample size information by making a judgment in line with

a few cases, or a small sample, that may run counter to a statistic based on many more instances. Because people so commonly receive contrasting information from different sources, this research has broad implications for situations in which people examine medical information, consumers compare products, and jurors weigh evidence. Educating students to consider sample size when comparing information continues to be important, in light of evidence suggesting that the number of encounters matters.

#### AUTHOR NOTE

This project was supported by NSF Grants SES-03-25080 and SBE-06-24098 to the second author, by research funds from Rutgers and NSF Grant REC-9720410 to the third author, and by a Rutgers University excellence fellowship and an NSF graduate fellowship awarded to the first author. We thank Dan Gottlieb and Randy Gallistel for their comments regarding the research, and Arpit Arora for coding subjects' responses. Correspondence concerning this article should be addressed to N. A. Obrecht, Psychology Department, Rutgers University, 152 Frelinghuysen Road, Piscataway, NJ 08854-8020 (e-mail: natalie@ruccs.rutgers.edu).

#### REFERENCES

- BORGIDA, E., & NISBETT, R. E. (1977). The differential impact of abstract vs. concrete information on decisions. *Journal of Applied Social Psychology*, *7*, 258-271. doi:10.1111/j.1559-1816.1977.tb00750.x
- BRASE, G. L., & BARBEY, A. K. (2006). Mental representations of statistical information. In A. Columbus (Ed.), *Advances in psychology research* (Vol. 41, pp. 91-113). Hauppauge, NY: Nova Science.
- 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. doi:10.1016/0010-0277(95)00664-8
- DEHAENE, S., DEHAENE-LAMBERTZ, G., & COHEN, L. (1998). Abstract representations of number in the animal and human brain. *Trends in Neurosciences*, *21*, 355-361. doi:10.1016/S0166-2236(98)01263-6
- FAGERLIN, A., WANG, C., & UBEL, P. A. (2005). Reducing the influence of anecdotal reasoning on people's health care decisions: Is a picture worth a thousand statistics? *Medical Decision Making*, *25*, 398-405. doi:10.1177/0272989X05278931
- FONG, G. T., KRANTZ, D. H., & NISBETT, R. E. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive Psychology*, *18*, 253-292. doi:10.1016/0010-0285(86)90001-0
- FREDERICK, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*, 25-42. doi:10.1257/089533005775196732
- GELMAN, R. (1996). Domain specificity in cognitive development: Universals and nonuniversals. In M. Sabourin, F. I. M. Craik, & M. Robert (Eds.), *Advances in psychological science* (Vol. 2, pp. 567-579). Hove, U.K.: Taylor & Francis.
- GIGERENZER, G., & HOFFRAGE, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*, 684-704. doi:10.1037/0033-295X.102.4.684
- GOTTLIEB, D., WEISS, T., & CHAPMAN, G. B. (2007). The format in which uncertainty information is presented affects decision biases. *Psychological Science*, *18*, 240-246. doi:10.1111/j.1467-9280.2007.01883.x
- GRICE, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41-58). New York: Academic Press.
- HASHER, L., GOLDSTEIN, D., & TOPPINO, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning & Verbal Behavior*, *16*, 107-112. doi:10.1016/S0022-5371(77)80012-1
- HASHER, L., & ZACKS, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General*, *108*, 356-388. doi:10.1037/0096-3445.108.3.356
- HASHER, L., & ZACKS, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist*, *39*, 1372-1388. doi:10.1037/0003-066X.39.12.1372
- HASHER, L., ZACKS, R. T., ROSE, K. C., & SANFT, H. (1987). Truly incidental encoding of frequency information. *American Journal of Psychology*, *100*, 69-91. doi:10.2307/1422643
- HERTWIG, R., BARRON, G., WEBER, E. U., & EREV, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, *15*, 534-539. doi:10.1111/j.0956-7976.2004.00715.x
- HUREWITZ, F., GELMAN, R., & SCHNITZER, B. (2006). Sometimes area counts more than number. *Proceedings of the National Academy of Sciences*, *103*, 19599-19604. doi:10.1073/pnas.0609485103
- KOKIS, J. V., MACPHERSON, R., TOPLAK, M. E., WEST, R. F., & STANOVICH, K. E. (2002). Heuristic and analytic processing: Age trends and associations with cognitive ability and cognitive styles. *Journal of Experimental Child Psychology*, *83*, 26-52. doi:10.1016/S0022-0965(02)00121-2
- LEHMAN, D. R., LEMPERS, R. O., & NISBETT, R. E. (1988). The effects of graduate training on reasoning: Formal discipline and thinking about everyday-life events. *American Psychologist*, *43*, 431-442. doi:10.1037/0003-066X.43.6.431
- LIPKUS, I. M., SAMSA, G., & RIMER, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, *21*, 37-44. doi:10.1177/0272989X0102100105
- LOEWENSTEIN, G. F., WEBER, E. U., HSEE, C. K., & WELCH, N. (2001). Risk as feelings. *Psychological Bulletin*, *127*, 267-286. doi:10.1037/0033-2909.127.2.267
- NISBETT, R. E., KRANTZ, D. H., JEPSON, C., & KUNDA, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review*, *90*, 339-363. doi:10.1037/0033-295X.90.4.339
- NISBETT, R. E., & ROSS, L. D. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice Hall.
- OBRECHT, N. A., CHAPMAN, G. B., & GELMAN, R. (2007). Intuitive *t* tests: Lay use of statistical information. *Psychonomic Bulletin & Review*, *14*, 1147-1152.
- OBRECHT, N. A., CHAPMAN, G. B., & SUAREZ, M. T. (2009). *Matching data with context: The integration of statistical variance and real-world knowledge*. Manuscript submitted for publication.
- PETERS, E., VÄSTFJÄLL, D., SLOVIC, P., MERTZ, C. K., MAZZOCCO, K., & DICKERT, S. (2006). Numeracy and decision making. *Psychological Science*, *17*, 408-414. doi:10.1111/j.1467-9280.2006.01720.x
- REYNA, V. F., & BRAINERD, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning & Individual Differences*, *7*, 1-75. doi:10.1016/1041-6080%2895%2990031-4
- SEDLMEIER, P., & GIGERENZER, G. (1997). Intuitions about sample size: The empirical law of large numbers. *Journal of Behavioral Decision Making*, *10*, 33-51. doi:10.1002/(SICI)1099-0771(199703)10:1<33::AID-BDM244>3.3.CO;2-Y
- STANOVICH, K. E. (2001). *How to think straight about psychology* (6th ed.). Needham Heights, MA: Allyn & Bacon.
- UBEL, P. A., JEPSON, C., & BARON, J. (2001). The inclusion of patient testimonials in decision aids: Effects on treatment choices. *Medical Decision Making*, *21*, 60-68.
- ZACKS, R. T., HASHER, L., & SANFT, H. (1982). Automatic encoding of event frequency: Further findings. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *8*, 106-116. doi:10.1037/0278-7393.8.2.106

#### NOTES

1. Latin-square version and presentation order of the 0-individual condition [ $F(4,155) = 5.32, p = .0005, \eta_p^2 = .1208$ ]; Latin-square version, presentation order of the 0-individual condition, presentation order of the 4- and 16-individual conditions [ $F(4,155) = 3.00, p = .0202, \eta_p^2 = .0719$ ]; Latin-square version, presentation order of the 4- and 16-individual conditions, and presentation order of the serial and summary conditions [ $F(4,155) = 3.10, p = .0172, \eta_p^2 = .0742$ ]; presentation order of the 0-individual condition, presentation order of the 4- and 16-individual conditions, and presentation order of the serial and summary conditions [ $F(4,155) = 4.34, p = .0389, \eta_p^2 = .0272$ ].
2. This interaction was not significant for the rating scale dependent measure, so we draw only partial support for our hypothesis.
3. There was a three-way interaction among probability-sample-size pairing, number range, and the order of data [ $F(2,180) = 3.25, p = .0410, \eta_p^2 = .0349$ ]. Story order and data order interacted [ $F(1,90) = 4.76, p < .0318, \eta_p^2 = .0502$ ].

(Manuscript received April 21, 2008;  
revision accepted for publication January 26, 2009.)