

Which Statistical Formats Facilitate What Decisions? The Perception and Influence of Different Statistical Information Formats

GARY L. BRASE*

University of Sunderland, UK

ABSTRACT

Research suggesting both a superiority of frequencies over single-event probabilities and of smaller reference classes over very large reference classes have utilized tasks that vary in computational complexity. The present research sought to simply and directly evaluate if—apart from simplifying statistical inference tasks—frequencies and small reference classes are clearer than other formats. After eliminating possible computational confounds, simple frequencies (based on small reference classes, e.g. 1/3) and to some extent relative frequencies (percentages, e.g. 33%) were perceived as clearer than absolute frequencies (based on very large reference classes, e.g. 90 million Americans) and single-event probabilities (e.g. 0.33). Concurrently, these different formats were evaluated in terms of their relative influence. Absolute frequencies were relatively more persuasive for smaller magnitudes (e.g. 2.7 million) but less persuasive for larger magnitudes (e.g. 267 million), as compared to analogous presentations. Single-event probabilities were judged to minimize the significance of information. Copyright © 2002 John Wiley & Sons, Ltd.

KEY WORDS judgment under uncertainty; reasoning; probability judgment; stimulus ambiguity; problem solving; social influence; rationality

The following exchange recently occurred on a national radio program (National Public Radio, 13 June 2000):

- ‘In 1999, 3800 cases [were] filed before the EEOC regarding wage discrimination.’ (Judith Appelbaum, vice president and director of employment opportunities at the National Women’s Law Center)
- ‘But there are 60 million women in the workforce, that’s 0.006%, or to look at it another way, 1 in 15 000 women.’ (Diana Furchtgott-Roth, fellow at the American Enterprise Institute)

Considering the above transcript, it is not surprising that many people subscribe to the view once expressed by Disraeli that ‘There are three types of lies: lies, damned lies, and statistics’. Behind this dismissal, however, is a serious psychological issue: Are people good or bad at evaluating statistical information? Are people routinely over-influenced, and then at other times under-influenced, by the statistical information

* Correspondence to: Gary L. Brase, Sunderland Business School, Division of Psychology, University of Sunderland, St Peter’s Campus, St Peter’s Way, Sunderland, England SR6 0DD, UK. E-mail: gary.braser@sunderland.ac.uk

they receive? If so, why? Research on statistical judgments under uncertainty has historically had a 'pessimistic' answer to these questions (Matlin, 1994). Results supporting this position indicate an appalling lack of statistical insight and computational competency in humans, who fall back instead on simple biases and rule-of-thumb heuristics (Kahneman et al., 1981). Other, more recent, research has outlined a relatively 'optimistic' answer to the question of human statistical competency. Research supporting this view usually shows how people become much better at using statistical information when that information is presented in ways that reflect the nature of the real world and mesh with the design of the human cognitive architecture (e.g. Brase, Cosmides, & Tooby, 1998; Gigerenzer & Hoffrage, 1995; Hoffrage, Lindsey, Hertwig, & Gigerenzer, 2000).

While these two perspectives are not actually as dichotomous as they sometimes appear (see Samuels et al., 2001), there is a very real debate about whether normal human decision-making abilities are relatively competent or relatively incompetent. The idea that, under certain situations, statistical judgments under uncertainty may be accurate and well calibrated holds clear and important implications for both the judgment under uncertainty field and the reasons why people continue to perceive statistical information as clear, important, and a basis for making decisions. Furthermore, this issue is not only important for gaining an accurate depiction and understanding of human competencies, but also for understanding how those competencies can be tapped into for applied uses of statistical information. For instance, some classically successful product advertisements are simply statements of statistical information: A certain soap is known for being 99 & 44/100ths pure; 4 out of 5 dentists recommend a particular gum for their patients who chew gum. A pessimist might argue that humans are ignorant of their own incompetence, and hence witless pawns for statisticians and marketers. An optimist might instead say that statistics, when applied carefully and judiciously, are generally used in a competent and responsible manner.

HOW FREQUENCIES FIT WITH BAYESIAN REASONING

A good deal of this debate has centered on the ability of people to engage in Bayesian reasoning. Bayesian reasoning, in a general sense, involves deriving the posterior probability of an event, such as the probability that a person has cancer based on the results of a test relevant to detecting cancer. The classic pessimistic view about peoples' abilities in this area is that 'In his evaluation of evidence, man is apparently not a conservative Bayesian: he is not a Bayesian at all' (Kahneman & Tversky, 1972, p. 450). Yet, those conducting research out of the optimistic approach have found that presenting the information in statistical word problems as natural frequencies, rather than as single-event probabilities, can dramatically improve people's statistical reasoning capabilities. These findings have led to a direct challenge to the above statement:

It now appears that this conclusion was premature. Frequentist problems elicit Bayesian reasoning. This finding adds to the growing body of literature that shows that many cognitive biases in statistical reasoning disappear, and good statistical reasoning reappears, when problems are posed in frequentist terms. (Cosmides & Tooby, 1996, p. 62)

One result of this debate is that it has fruitfully uncovered a conceptual ambiguity regarding judgments under uncertainty. What exactly is 'Bayesian reasoning'? A more restricted answer, often favored by the pessimists, is that Bayesian reasoning is applying the usual formulation of Bayes' Theorem¹ to a problem in order to reach the correct answer. A more general answer is usually favored by the optimists: Bayesian reasoning is working through a problem in some way (using Bayes' Theorem in any formulation or using any other means) such that the correct answer is reached. In short, the restricted view of Bayesian reasoning requires

¹Bayes' Theorem is usually given as: $p(H|D) = \frac{p(D|H)p(H)}{[p(D|H)p(H) + p(D|\sim H)p(\sim H)]}$, where H refers to a hypothesis about the world and D refers to the data relevant to that hypothesis. The term for which the theorem is solved ($p(H|D)$) is called the posterior probability of the hypothesis.

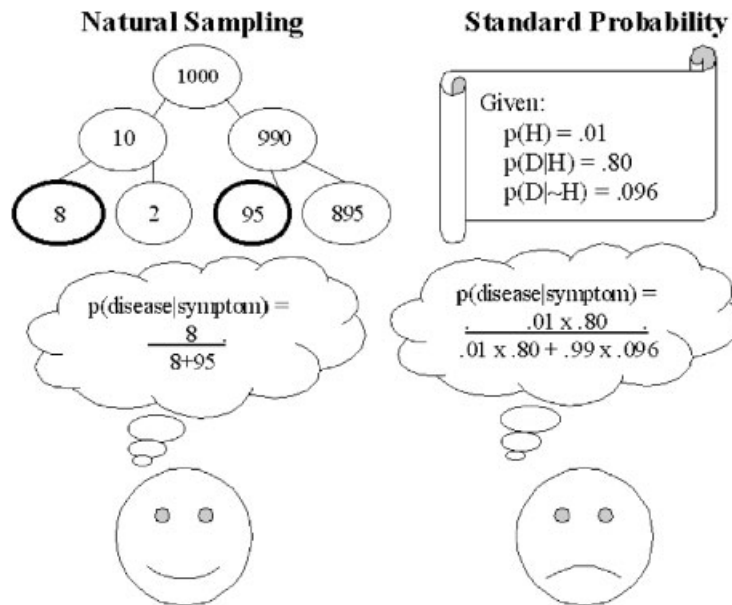


Exhibit 1. Differences in computational complexity in conducting Bayesian inferences, due to the information representation (natural sampling of frequencies versus single-event probabilities). (Adapted from Gigerenzer & Hoffrage, 1995)

using Bayes' Theorem; the general view of Bayesian reasoning requires reaching the correct answer (the posterior probability).

The reason this distinction is important is because, as Gigerenzer and Hoffrage (1995) have pointed out, information given as naturally sampled frequencies (hereafter, 'natural frequencies') does not require the use of Bayes' Theorem in order to calculate the posterior probability.² In other words, natural frequencies allow one to be generally Bayesian without being restrictedly Bayesian. More careful distinctions could resolve this aspect of the disagreement, but another aspect of the disagreement is more substantial. Besides not being restrictedly Bayesian, natural frequencies allow a way to reach the general Bayesian goal of the posterior probability with much *simpler* computations (see Exhibit 1). So the fact that natural frequencies facilitate Bayesian reasoning (in the general sense) can be argued as being due to both (a) some inherent superiority (i.e. cognitively privileged representational status) of frequencies, and (b) the computational simplification that results from using frequencies. In fact, no one debates that the latter is the case; natural frequencies allow a computationally simpler way to calculate a posterior probability (Gigerenzer & Hoffrage, 1995; Brase et al., 1998). What is less certain is the claim, by the optimists, that frequencies are more generally a privileged representational format as compared to other numerical formats.

ECOLOGICAL RATIONALITY AND COMPUTATIONAL SIMPLIFICATION

In typical judgments under uncertainty studies, participants are asked to answer a statistical word problem, often a problem that involves the application of Bayesian reasoning. A correct response is taken as evidence that the person successfully understood the problem and was able to work it through to an accurate conclusion. Answers that are wrong in some way indicate some form of conceptual or computational incompetence.

²Stated most simply, a natural sampling framework is an arrangement of frequency information of category instances into subset relationships. This structure has also been variously called the *subset principle*, *partitive formulation*, and *set inclusion mental models*.

The frequency assumption

Data presented as natural frequencies within Bayesian word problems generally lead to more correct answers (i.e. estimates of the posterior probability) than other data formats (e.g. single-event probabilities). The optimistic view places the credit for this facilitation on an idea recently termed *ecological rationality*:

Evolutionary theory asserts that the design of the mind and its environment evolve in tandem. Assume—*pace* [Steven J.] Gould—that humans have evolved cognitive algorithms that can perform statistical inferences. These algorithms, however, would not be tuned to probabilities or percentages as input format. . . . We assume that as humans evolved, the ‘natural’ format was *frequencies* [italics in original] as actually experienced in a series of events, rather than probabilities or percentages. (Gigerenzer & Hoffrage, 1995, p. 686)

The approach delineated by ecological rationality is essentially a model of bounded rationality, as outlined by Herb Simon (1956, 1990), but with a renewed focus on the recurrent structures of the environment as important elements in forming criteria of human rationality (for further discussion, see Gigerenzer et al., 1999). Briefly, those features of the world that have been stable or recurrent during evolutionary history are features that the human mind has been able to capitalize on in developing judgment and decision-making abilities that are superior to any similar abilities that ignored such features (see also Shepard, 1984, 1987, for a similar perspective within the domain of perception). Tasks that fail to resemble the structure of how information exists in the real world often fail to evoke these abilities.

As a result, the ecological perspective provides *three interrelated* reasons why frequencies—and specifically natural frequencies—would predominantly be the presentation mode that the mind uses in making judgments about the environment (i.e. be privileged). First, information about one’s observations of the world (i.e. counts of things that either happen or do not happen) are most directly represented as frequencies (also see converging conclusion in Harries & Harvey, 2000). Second, as such frequency information is acquired it can be easily, immediately, and usefully incorporated with past frequency information via the use of *natural sampling*, which is the method of counting occurrences of events as they are encountered and storing the resulting knowledge base for possible use later (hence producing natural frequencies; on natural sampling see Kleiter, 1994; Aitchison & Dunsmore, 1975). Third, data in natural frequencies retains more valuable information that is lost in other formats. Specifically, sample sizes are retained and usable (e.g. a rate of 20% (or 0.20) could be based on either a sample of 1 out of 5 or a sample of 10 000 out of 50 000), the knowledge database as a whole is easily updated with new information, and the original data can be reorganized to construct novel reference classes if necessary (see Brase et al., 1998, for further elaboration on these reasons).

For all these above reasons, it has been hypothesized in some recent research (Brase et al., 1998; Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995), that the mind has over evolutionary time developed abilities specifically designed to make informed statistical judgments based on natural frequencies.

The pessimistic view disagrees, and instead places the credit for the facilitative effect of natural frequencies entirely on the computational simplification that results from the use of frequencies:

A more accurate conclusion is that their respondents are competent at whole number arithmetic, which is anyway hardly surprising in view of the fact that they are often university students. But with probabilistic reasoning, and especially with reasoning about frequency probabilities, Cosmides and Tooby’s results have very little to do at all, despite their dramatic claims. (Howson & Urbach, 1993, p. 422)

The dilemma so far has been that little data exists to disentangle these two possibilities.³ Posing a Bayesian reasoning problem in natural frequencies simultaneously changes the numerical format and simplifies the

³Gigerenzer and Hoffrage (1995, 1999) eliminated most of the differences in computational complexity between isomorphic problems expressed in natural frequencies and probabilities (percentages) and found that natural frequencies still produced superior judgments. The probability problem, however, still required the combining of two probabilities (e.g. 0.8% of 10.3%) to provide an answer, whereas the frequency version required only reporting the frequencies (e.g. 8 out of 103). One could argue that this difference in computational complexity was still a significant obstacle.

required computations. One objective of this research project is to determine if frequencies are clearer and easier to understand (as the optimists claim), aside from the computational ease they also can provide. I will refer to this claim as the *frequency assumption*—which is an assumption that the optimists make and the pessimists do not make.

It may seem odd, but the relative clarity and persuasive effectiveness of various statistical presentation formats has not been evaluated directly. Certain formats have been found to facilitate accurate judgments under uncertainty, and the inference has been drawn that those formats are clearer. But as recently stated by Johnson-Laird et al. (1999), 'It is indeed crucial to show that the difference between frequencies and probabilities transcends mere difficulties in calculation and that difficulties with problems about unique events are not attributable merely to difficulties in numerical calculation' (p. 81). Girotto and Gonzalez (2001) have shown good judgment performances with information given in 'subset relations' that are isomorphic with a natural sampling structure and 'chances' that are isomorphic with frequencies (e.g. '5 chances out of 100'), but these results can be argued as simply being due to the use of natural frequencies thinly disguised and relabeled as something different. Evans, Handley, Perham, Over, and Thompson (2000) engage in a similar exercise in which natural sampling structure is redubbed as 'a set inclusion mental model' and Macchi (2000) takes the same approach by calling the natural sampling framework a 'partitive formulation' (see Hoffrage et al., in press, for elaboration on this view).

The simplest and most direct initial answer to this question, although no longer an assay of Bayesian reasoning ability, is to simply ask people which numerical format is clearest and easiest to understand. One aspect of the present research is to do just that.

The reference class assumption

Another facet of the ecological rationality viewpoint, also relevant to this study, deals with the reference class of the groups or categories that one is reasoning about. Research by X. T. Wang (1996a–c; Wang & Johnston, 1995; see also Fetherstonhaugh et al., 1998) has documented that in making judgments about populations of people—especially life-and-death decisions—people are sensitive to population sizes. Specifically, when people are asked to make decisions about a small-group or family sized population (under 100 people) the traditionally observed framing effects (Tversky & Kahneman, 1981) of decision malleability disappear. The explanation for this pattern of sudden consistency stems from the rationale that these smaller population sizes are on scales of magnitude that humans have directly and recurrently dealt with over evolutionary history. Unrelated research tangentially supports the idea that human judgment is both most accurate and most stable when dealing with information on scales that are more consonant with human experiences, both ontogenetically and phylogenetically (Freudenburg & Rursch, 1994; Thompson & Schumann, 1987).

What is important for the present purposes is that this view suggests very large reference classes are not perceived veridically, but rather in a manner somewhat similar to the changes in difference thresholds in sensory perception. Specifically, two items on an extremely large scale of magnitude (e.g. 90 million and 180 million) are not easily distinguished from one another, whereas two items on a much smaller scale (e.g. 33% and 66%) are much more easily discriminated from one another. Very large numbers, such as those in the millions or billions, will be perceived as uniformly large, and will not necessarily be perceived as accurately in relation to each other. As with the frequency assumption, however, an argument could be made that this *reference class assumption* makes the same inferential jump from performances on complex statistical reasoning problems to the theoretical claim that smaller reference classes are fundamentally more ecologically rational than large reference classes (i.e. mathematical calculations are typically easier when using smaller numbers). A second aspect of the present research is to assess this reference class assumption.

APPLIED USES OF STATISTICAL FORMATS

A final reason to explore this issue is the potential for applied uses. Are there ways of presenting data that are more easily digested by consumers? Similarly, are there also ways of presenting data that may obfuscate matters to suit the motives of those who supply the data? The details of answers to these questions will differ depending on whether one asks a pessimist or an optimist.

Previous applied research has found results that are taken as generally fitting with the pessimistic view. These include studies conducted within applied domains such as jury decision making (Thompson & Schumann, 1987), understanding oral contraceptive risk (Halpern, Blackman, & Salzman, 1989), public school administration (Brown & Newman, 1982), and evaluating environmental hazards (Freudenburg and Rursch, 1994). All these studies note both an error-proneness and a malleability in subjects' responses based on the presented information, and focus on strategies to minimize the damage of these bad judgment abilities (e.g. by avoiding reliance on human judgments entirely). The optimistic viewpoint has only recently begun to explore the implications for applied fields of their insights into data presentation. Recent findings, however, suggest that these implications may be substantial (Hoffrage et al., 2000). Gigerenzer, Hoffrage, and Ebert (1998), for example, found that the large majority of AIDS counselors dealing with a low-risk client in their study (a non-drug-using heterosexual male) were unable to communicate the meaning of an AIDS test result in an accurate and intelligible manner. These authors, however, were able to construct simple and concise responses that an 'ideal' counselor could have given by using natural frequencies. In sum, choices about the method of presentation used for statistical information may benefit from knowledge of the principles of ecological rationality.

PREDICTIONS

An optimistic view of human decision making indicates that gaining an accurate understanding of information almost always benefits from the information being presented in ecologically rational ways (e.g. natural frequencies and small reference classes). This assertion, however, has not been tested directly, but rather inferred from statistical reasoning task performances. At issue is not only a theoretical question of how the mind is designed to process statistical information, but also the resulting implications for applying such theoretical knowledge to real-world concerns. Specifically, two aspects of how statistical information is perceived can be assessed: (1) the understandability, or clarity, of the information to recipients, and (2) the impact of the information as applied to some topic; how much the information influences the attitudes and behaviors of recipients. Two general hypotheses were advanced, based on the principles of ecological rationality:

- (1) Information presented as small-scale, frequencies (i.e. simple frequencies and relative frequencies (percentages)) will be *more easily* understood than other formats, regardless of the magnitude of the numbers used (e.g. 1 in 100, 1 in 3, 2 in 3, or 99 in 100). On the other hand, information presented in single-event probability formats (e.g. 0.01, 0.33, 0.67, or 0.99) will be *less easily* understood than other formats (simple frequencies, relative frequencies (percentages), or absolute frequencies in the US population), regardless of the magnitude of the information used.
- (2) In terms of persuasive impact on both attitudes and potential behaviors (i.e. social influence), large-scale, absolute frequency information (for the US population: 0.01 = 2.7 million, 0.33 = 90 million, 0.66 = 180 million, and 0.99 = 263 million) will be perceived as relatively similar in influence across different magnitudes, as compared to other formats (e.g. simple frequencies, relative frequencies, and single-event probabilities). Of course, one expects that larger magnitudes will have more impact (e.g. a disease affecting 66% of people is more impressive than one affecting 1% of people). In line with the above discussion regarding the reference class assumption, however, large-scale absolute frequencies

will have proportionately greater impact when referring to small magnitudes (e.g. 0.01), as compared to other information formats (because they will be perceived as simply a 'very large' quantity). Conversely, large-scale, absolute frequency information will have proportionately less impact when referring to large magnitudes (e.g. 0.99), as compared to other information formats, because although these are very large quantities their proportional magnitude will not be well understood. In summary, an interaction should occur such that large-scale, absolute frequency information will be both more influential in decisions (with small magnitudes) and less influential in decisions (with large magnitudes), as compared to other presentation formats.

EXPERIMENT 1

Method

Participants

Participants were 120 undergraduates at a mid-sized American university (74 females and 46 males). All consented to participate in exchange for extra credit in psychology classes. Their average age was 20.8 years.

Materials and procedure

Each participant was given a booklet that consisted of a brief instruction section followed by the actual study materials. The materials consisted of four items of statistical information, each of different magnitudes, and each of which was couched within a context story relevant to an applied area of psychology that commonly makes use of statistical information (disease prevalence, education, marketing, and drug efficacy; see Appendix). The format of the statistical information was manipulated between participants, using simple frequencies (e.g. 1 in 3), single-event probabilities (e.g. 0.33), relative frequencies (e.g. 33%), and absolute frequencies for the US population (e.g. 90 million). Out of the 16 conditions (4 quantities/stories \times 4 statistical formats), each participant received just one version of each story and format, with each item on its own sheet. Both the order of the four items and the orders in which the different types of statistical information were presented were counterbalanced to avoid order effects.⁴

Each of the items was followed by three probe questions that asked participants to evaluate the clarity, impact, and monetary pull of the statistical information. For the questions regarding clarity and impact, responses were collected on a 5-point scale with verbal labels (e.g. Somewhat unclear, Fairly clear, Clear, Very clear, Extremely clear). For the question regarding monetary pull, responses were collected by asking participants to allocate an amount out of every \$100 dedicated to the particular issue addressed by the statistical information. There was no time limit for completing this study, and all participants finished within 15 minutes.

Results

Which frequencies are judged to be clearer than single-event probabilities?

Exhibit 2 shows the mean ratings of how clear and easy to understand each statement was, across both the different statistical formats and the different magnitudes/context stories. The predictions regarding these data are hypothesized to occur independent of the context story and the magnitude of the numerical information used. In other words, any effects should involve just the statistical formats, and not statistical magnitudes or context stories. One-way independent samples analyses of variance confirmed that there

⁴One possible limitation of the present study is that the levels of statistical quantity are not varied independently of the different context stories. The responses of the participants indicate that this potential confound probably did not affect the ratings of the participants (i.e. clarity ratings were unrelated to the quantity/context story, and importance ratings increased in a general linear manner as a function of the quantity), but a future study (with four times as many subjects) could counterbalance these factors.

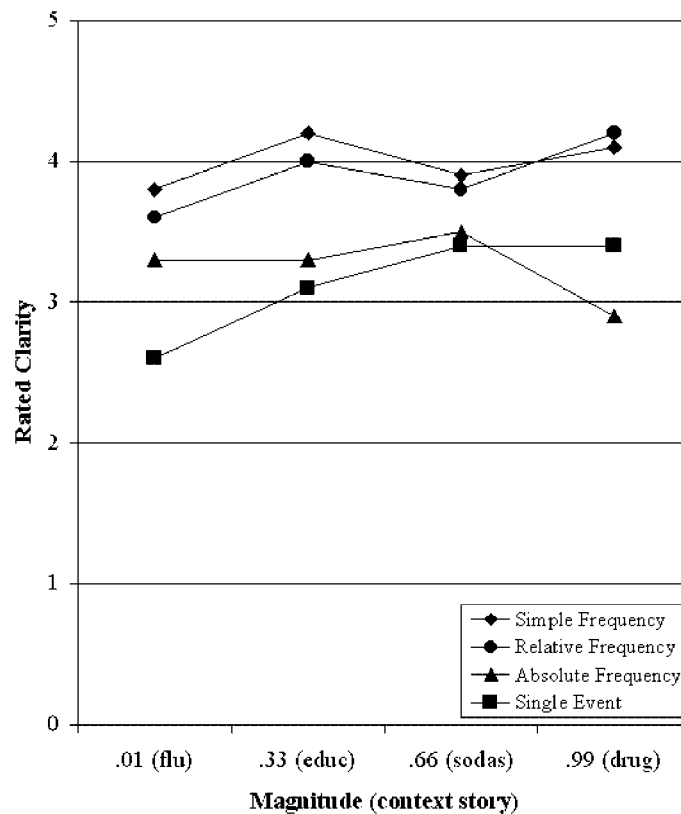


Exhibit 2. Mean ratings of statistical information presented in different formats, regarding the extent they are clear and easy to understand, presented across different magnitudes (context story)

were no differences in clarity ratings within each statistical format type, comparing across the context stories/statistical magnitudes (single event probabilities: $F(3, 116) = 2.14$, $p = 0.10$; simple frequencies: $F(3, 116) = 0.977$, $p = 0.41$; relative frequencies: $F(3, 116) = 1.50$, $p = 0.22$; absolute frequencies: $F(3, 116) = 1.07$, $p = 0.36$).

Collapsing across these conditions, however, reveals significant differences in rated clarity between the different types of statistical formats ($F(3, 476) = 15.42$, $p < 0.001$, $\eta^2 = 0.09$). The mean ratings for the collapsed conditions are: 3.13 for single-event probabilities, 3.24 for the absolute frequencies, 3.89 for the relative frequencies (percents), and 3.98 for the simple frequencies. Scheffé *post hoc* analyses indicate that, as predicted, both simple frequencies and relative frequencies are seen as significantly clearer than both single-event and absolute frequency information.

Does the impact of absolute frequencies change with magnitude?

The second prediction is addressed by the second and third questions posed to participants. Common sense indicates that increasing magnitudes, regardless of format, should lead to correspondingly greater impact. However, this pattern is predicted to be relatively attenuated at both extremes for information given in absolute frequencies. Specifically, the mean ratings of impact (as compared to other formats) should be significantly higher for absolute frequency information regarding very low magnitudes, but significantly lower for absolute frequency information regarding very high magnitudes.

Exhibits 3(a) and 3(b) show the mean ratings of how impressive each statement was (a) and how much monetary pull each statement had (b), across both the different statistical information formats and the different context stories/quantities. The greater the statistical quantity was in a statement, the more impressive and eliciting of money it was (a pattern that may well be significant, but is irrelevant to the present research). The patterns of interest here are the relative ratings of the different statistical formats within each statistical quantity. As predicted, at the lowest magnitude (0.01) there are significant differences ($F(3, 116) = 6.78, p < 0.001, \eta^2 = 0.15$) in impressiveness ratings. Scheffé *post hoc* tests showed that both simple frequencies and absolute frequencies were significantly *more impressive* than both the single-event probabilities and relative frequencies. At the two intermediate statistical magnitudes (0.33 and 0.66) there were no significant differences between any of the presentation formats (0.33 format: $F(3, 116) = 1.61, p = 0.19$; 0.66 format: $F(3, 116) = 1.13, p = 0.34$). At the highest magnitude (0.99), significant differences again appeared ($F(3, 116) = 6.25, p < 0.001, \eta^2 = 0.14$), and Scheffé *post hoc* tests concluded that absolute frequencies were rated significantly *less impressive* than both the simple frequencies and relative frequencies.

For the monetary pull measure, none of the statistical magnitude conditions showed significant differences (0.01 level: $F(3, 116) = 0.847, p = 0.47$; 0.33 level: $F(3, 116) = 2.69, p = 0.07$; 0.66 level: $F(3, 116) = 0.703,$

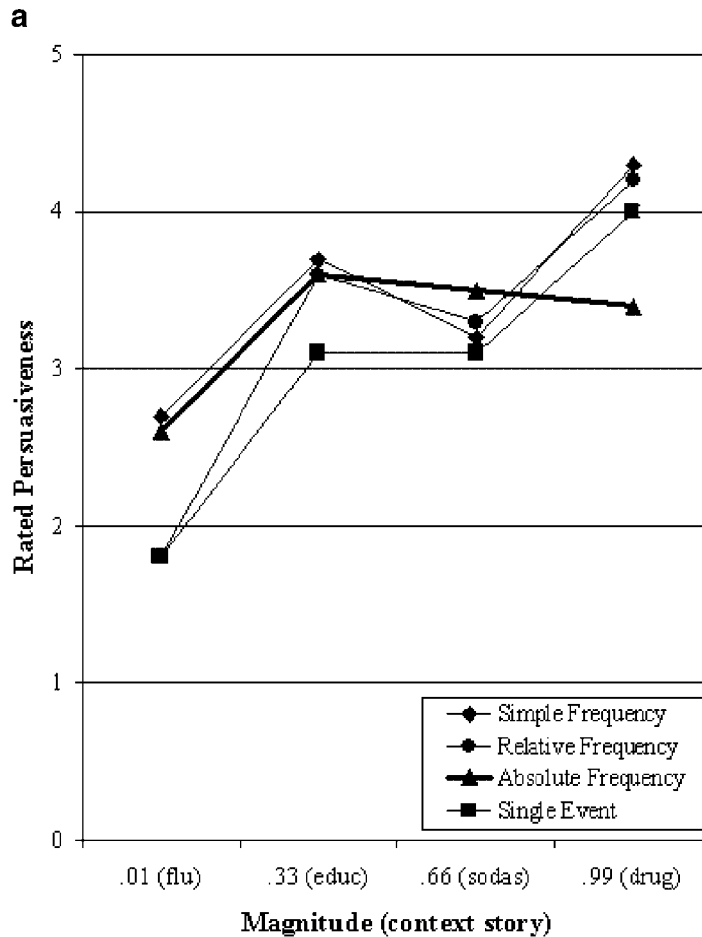


Exhibit 3. Mean ratings of statistical information presented in different formats, regarding the extent they are impressive (a) and elicit money (b), presented across different magnitudes (context story)

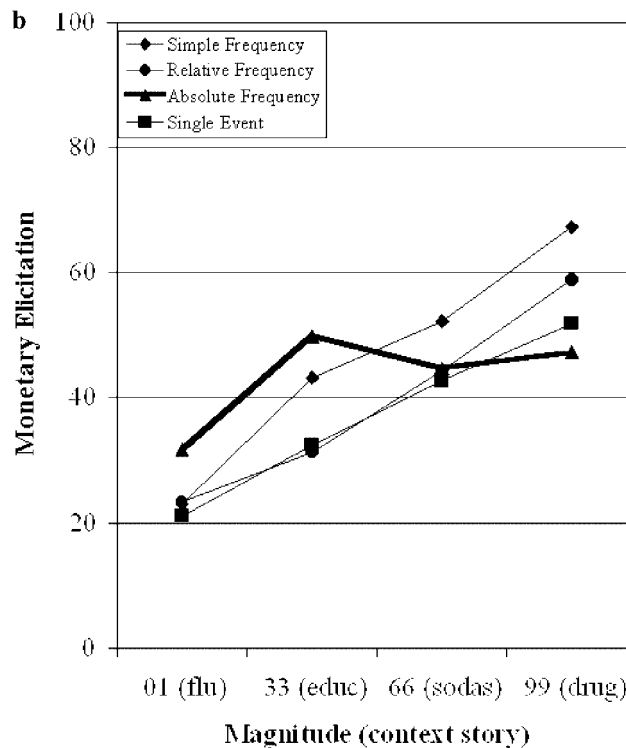


Exhibit 3. Continued

$p = 0.55$; 0.99 level: $F(3, 116) = 2.23$, $p = 0.09$). There are at least two possible reasons for this difference between the findings of general impressiveness and monetary pull (which should be related to some extent). First, these measures were on different scales. Whereas impressiveness was measured on a 5-point rating scale, monetary pull was measured on a 100-point scale (in dollars), and this may have had some effect on participants' ratings. Second, there could be an order effect such that the second of these items was reacted to differently. Nevertheless, a comparison of Exhibits 3(a) and 3(b) suggests that both of these items produced similar relationships between the impact of different statistical formats; specifically, absolute frequencies appear to have a relatively exaggerated impact at very low magnitudes and a relatively muted impact at very high magnitudes.

EXPERIMENT 2

A possible confound in Experiment 1 is the fact that the absolute frequency condition did not contain the appropriate reference class size information. In fact, the actual population of the United States is very rarely given alongside statistical information about how many US citizens do, believe, or have some particular thing. So whereas the absolute frequency stimuli in Experiment 1 certainly are ecologically valid, they may be considered inconsistent with the other stimuli. Relative frequencies (i.e., percentages) and single-event probabilities are normalized to set scales and thus are not influenced by this consideration. Simple frequencies (e.g. 1 out of 3) explicitly give a reference class, although it is generally not the true size of the reference class but rather a reduced version. How well people comprehend the actual size of the reference class could be a factor in the clarity of this presentation format. Experiment 2 was designed to replicate the

absolute frequency condition, but with a reference class (270 million) included in the information given to the participants.

Method

Participants

Participants were 122 undergraduates at a mid-sized American university (80 females and 42 males). All consented to participate as part of a classroom demonstration. Their average age was 23.7 years.

Materials and procedure

Each participant was given a sheet that included a brief instruction section followed by the actual stimuli materials. The materials for each participant consisted of one of the four items of absolute frequency information, each of different magnitudes and with different context stories, as used in Experiment 1, but with an appended statement in parentheses: '(The population of the United States of America is about 270 million people).' The statistical information item was followed by the same three probe questions as used in Experiment 1 (evaluating the clarity, impact, and monetary pull of the statistical information).

Results

The addition of the reference class information had no effect on the rated clarity of the statistical information, as compared to the analogous conditions in Experiment 1 ($F(1, 134) = 0.007, p = 0.93$). In fact, on the 5-point rating scale used to evaluate clarity, the largest difference observed between experiments for any of the magnitudes was one-third of one point. Similarly, there was no significant difference in comparing the analogous results for the 5-point scale rating of the impact of the information ($F(1, 134) = 0.538, p = 0.46$), with the largest difference observed for any of the conditions being 7/10ths of a point on the 5-point scale. The monetary pull probe question showed a significant difference between the present results and the results of Experiment 1, with the responses in this second experiment being overall lower (i.e. being overall less elicited of money: means at 0.01 level: \$18.44, at 0.33 level: \$20.60, at 0.66 level: \$41.76, at 0.99 level: \$38.87; $F(1, 134) = 13.04, p < 0.001$). What this would mean in comparison with the responses to the other statistical formats used in Experiment 1 is that the absolute frequency results would be more similar to the other format at the 0.01, 0.33, and 0.66 levels, but more dissimilar (in the predicted direction) at the 0.99 level. In other words, the pattern obtained in Experiment 2 continues to indicate that very large absolute frequencies are distinctly ineffective in soliciting proportionately greater monetary investment.

EXPERIMENT 3

Both Experiments 1 and 2 collected participant responses with likert-type scales, which are commonly used to measure attitudes and reactions to information or situations. It would be useful, however to have converging results using a somewhat different methodology, as well as different actual stimuli contents. Experiments 3 and 4 were designed to seek those results, using a methodology somewhat closer to those in typical decision making tasks (i.e. choices between alternatives).

Method

Participants

Participants were 86 undergraduates at a northern England university (60 females and 26 males). All consented to participate as part of a classroom demonstration. Their average age was 24.2 years.

Materials and procedure

Each participant was given a sheet that requested their age and sex, followed by a task description:

A variety of harmful bacteria (such as *Campylobacter*, *Salmonella*, and *E. coli*) can exist in different food products. Normal cooking processes generally kill all of these bacteria, although human infections do sometimes occur due to undercooking. Below are the rates of occurrence of these bacteria in different foods (*prior* to cooking).

Three items followed, each referring to a different type of meat (pork, beef, or chicken), and each expressing the associated bacteria rate in a different statistical format (all rates roughly equivalent to 20%). The order in which the meats occurred was counterbalanced, and the order of the statistical formats was reversed for half the participants (i.e. given in the order below or the reverse):

- A. There is a 0.20 probability that a piece of [pork/beef/chicken] is infected
- B. [990/2250/1487] million pieces of [chicken/pork/beef] in the UK are infected each year
- C. 1 out of every 5 pieces of [beef/chicken/pork] are infected

Three questions regarding these items were then asked:

- (1) Which of the above statements gives you the *best* understanding of the situation?
- (2) Which of the above statements sounds like the *most* serious problem?
- (3) Which of the above statements sounds like the *least* serious problem?

Results

Responses to each of the three questions were coded by statistical format and collapsed across the different contents (types of meat). Non-standard responses (e.g. more than one answer for a given item or 'depends') were discarded. The results were very clear: 86% of participants (73 out of 85) concluded that the simple frequency information provides the best understanding of the situation, 78% (66 out of 85) decided that absolute frequencies sound like the most serious problem, and 87% (72 out of 83) judged the single-event probabilities as sounding like the least serious problem. Each of these are significant patterns ($\chi^2(2) = 105.91$, $p < 0.001$; $\chi^2(2) = 79.08$, $p < 0.001$; and $\chi^2(2) = 106.72$, $p < 0.001$, respectively) (Exhibit 4).

EXPERIMENT 4

The objective of Experiment 4 was to replicate the results of Experiment 3 as well as extend them by adding relative frequency (percentages) information as a condition and using yet another type of context for the statistical information.

Method*Participants*

Participants were 73 undergraduates at a northern England university (59 females and 14 males). All consented to participate as part of a classroom demonstration. Their average age was 22.7 years.

Materials and procedure

Each participant was given a sheet that requested their age and sex, followed by a task description:

The following statements, all revealing the same information, were made by different people regarding wage discrimination against women in the workplace.

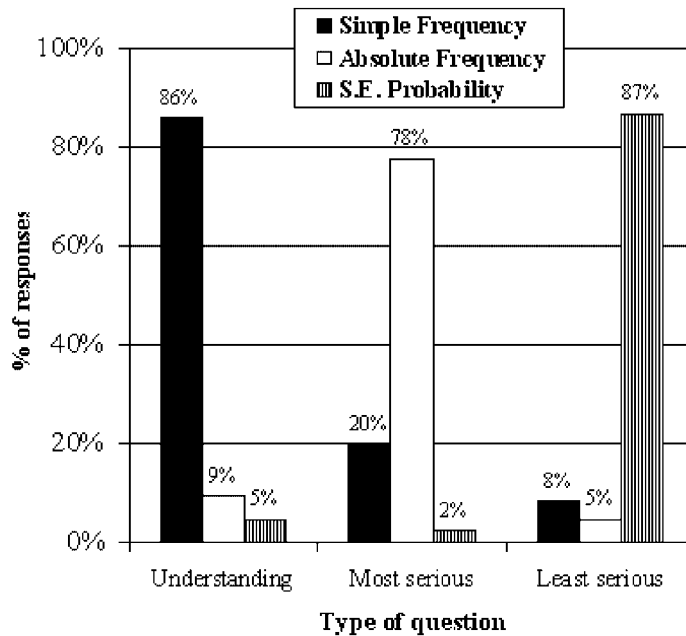


Exhibit 4. Percentages of participants in Experiment 3 that indicated each of the statistical formats as being easiest to understand, sounding like the most serious problem, and sounding like the least serious problem

Four items followed, each referring to a different person as a source and each expressing the same information expressed in different statistical formats. The order in which the items occurred was counterbalanced:

- Angela: 'In 1999, 3800 complaints were filed by women in the workforce regarding wage discrimination.'
- Betty: 'In 1999, 0.006% of the 60 million women in the workforce filed complaints regarding wage discrimination.'
- Crystal: 'In 1999, 1 in every 15 000 women in the workforce filed complaints regarding wage discrimination.'
- Diane: 'In 1999, there was a 0.006 probability⁵ that a woman employee would file a complaint regarding wage discrimination.'

Three questions regarding these items were then asked:

- (1) Who is trying to exaggerate the situation the most?
- (2) Who is trying to minimize the situation the most?
- (3) Who is being the clearest?

Results

The results, while not as dramatic as those in Experiment 3, are still very clear: 48% of participants conclude that the simple frequency information is the clearest (followed by 34% choosing the absolute frequencies as clearest), 58% of participants decided that absolute frequencies exaggerated the situation the most, and 70%

⁵This probability, to be consistent with the other statements should actually be 0.00006, but it was decided that such a number of zeros could bias potential results. The actual findings of this experiment are actually more notable for the fact that $p = (0.006)$ is greater than 0.006%.

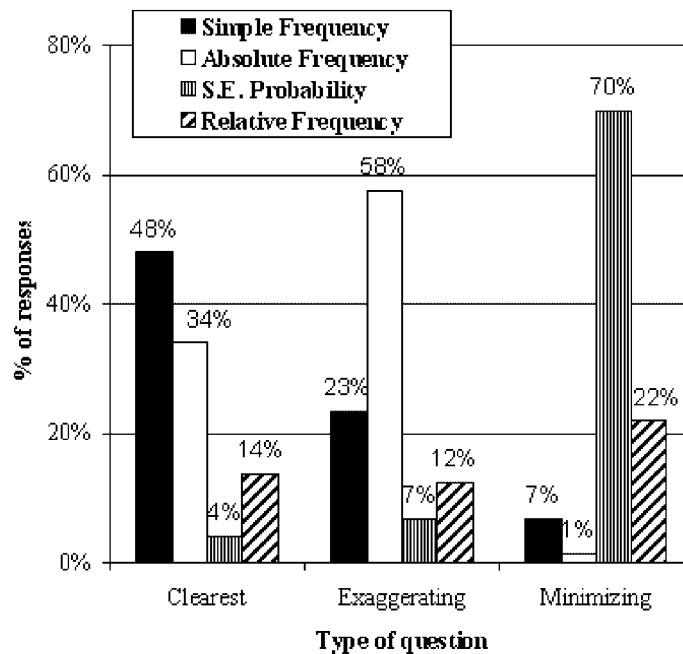


Exhibit 5. Percentages of participants in Experiment 4 that indicated each of the statistical formats as being clearest, most likely to be used to exaggerate the situation, and most likely to be used to minimize the situation

of participants judged the single event probabilities as minimizing the situation the most. Each of these are significant patterns ($\chi^2(3) = 45.30$, $p < 0.001$; $\chi^2(3) = 84.97$, $p < 0.001$; and $\chi^2(3) = 34.34$, $p < 0.001$, respectively) (Exhibit 5).

DISCUSSION

These results of these experiments support the predictions regarding statistical information formats. Across all studies, statistical information presented as simple frequencies is viewed as being clearer and easier to understand than the same information presented in other types of representational formats, and information presented as single-event probabilities is particularly difficult to understand. These findings, which hold across varying magnitudes and across a variety of contexts, conform with a number of recent findings in the judgment under uncertainty literature (e.g. Gigerenzer, 1991; Cosmides & Tooby, 1996; Brase et al., 1998). Furthermore, these results indicate that frequencies actually are, apart from their general simplification of statistical reasoning task complexity, facilitative of peoples' understanding of data.

Regarding the amount of impact statistical information has on people, the presentation format (specifically, absolute frequencies versus other formats) was found to have a significant effect that interacts with the relative magnitude of the information. At low magnitudes (0.01), absolute frequency information has as much or more impact on participants than other presentation formats. Intermediate magnitudes (0.33 and 0.66) yielded no significant differences between the various formats, but at high magnitudes (0.99), information in an absolute frequency format had less impact than other formats. Furthermore, this general pattern of results replicates from general ratings of impressiveness to participants' allocations of monetary amounts based on the data (albeit not with statistical significance). The more generally flattened level of impact for absolute frequencies as the magnitude of the numbers increased (both for general impressiveness

and for eliciting money) was also found to hold regardless of whether the reference class for the numbers was given or not. Finally, absolute frequencies are consistently recognized as the format of choice for increasing or exaggerating the importance of statistical information using fairly low magnitudes (20% in Experiment 3 and 0.006% in Experiment 4), whereas single-event probabilities are consistently recognized as the format of choice for decreasing or minimizing the importance of statistical information.

Not all frequencies are equal

There is some incongruity between what people find most clear and easy to understand when it comes to statistical formats and what is theoretically the most useful statistical format. Simple frequencies—and to a somewhat lesser extent relative frequencies—are most clear, but are also normalized (i.e. they do not preserve the true sample size) and therefore do not communicate potentially important information. This means that these formats can actually be deceptive (in terms of obfuscating the sample size upon which they are based) at the same time as they are perceived as being clearest. Absolute frequencies carry the most information, as they retain both sample size information *and* retain the ability to form an accurate natural sampling structure, however absolute frequencies become unwieldy with very large samples. It may be the case that there is a trade-off, in terms of cognitive economics, between the benefits of actively retaining absolute frequency information and the costs of keeping such information as it becomes very extensive. How exactly this trade-off between types of information representations is implemented remains to be explored (particularly issues regarding if and how the representation of the numerical information changes from relatively direct, quantified forms (explicit representation) to presumably a more intuitive and qualitative forms (implicit representation). It is interesting to note that absolute frequency information was found to be a relatively clear presentation format in Experiment 4 (by 34% of participants), and the absolute frequency information in this experiment (3800) was the smallest magnitude tested (in simple frequency terms, 1 in 15 000). This may indicate a general point at which absolute frequency information begins to be tractable in terms of human abilities to understand and work with statistical information.

In considering the contrast between information presented as frequencies versus as probabilities, it might be argued that frequencies are easier to understand simply because they are more prevalent and more often observed in the natural world. A circular argument seems to arise in which our evolved dispositions lead us to prefer frequencies (we are more comfortable with using them), so we use frequencies more often, so frequencies are more common in the world, so we are more comfortable with frequencies. This circularity can even be used on the present results on the clarity and understandability of different statistical formats; are they naturally clear and understandable or do we only have a metacognitive awareness of our relative comfort/familiarity with different statistical formats? Research from developmental psychology suggests that this circular pattern, although it may have some influence, is not a complete explanation. Young children develop very characteristic areas of mathematical understanding as a reliably developing aspect of normal human development, to the extent that these abilities are sometimes called 'intuitive mathematics' (e.g. Brase, in press; Gallistel & Gelman, 1992; Geary, 1995; Gelman, 1998; Hartnett & Gelman, 1998; Wynn, 1992, 1995, 1998a,b). The types of abilities that emerge within this intuitive mathematics are all based in frequency representations (e.g. numerosity, ordinality, and whole number addition). Thus, before a child has learned to recognize the physical representations—or even the verbal names—for numbers, frequencies are already clearly set as the primary representational format.

Making statistics clearer

It is important to point out that the results given here provide just one converging line of evidence that frequencies are preferred as a presentation format for information in the world and that they are more easily used by the human mind than other formats. The present data consists of self-reports and forced alternative

decision making items regarding clarity and social influence, which by themselves might possibly be contested by questioning the veracity or insightfulness of the research participants. If, however, one compares these results to the finding of previous studies using statistical reasoning problems, it seems clear that these results fit quite well with actual performances using frequencies and single-event probabilities. Simple frequencies are evaluated as being clearer and easier to understand than other formats, and natural frequencies that are placed on scales similar to these simple frequencies (i.e. involving hundreds or thousands of items) also elicit statistical reasoning performances that are superior to performances with other formats. Multiple studies, using different methodologies, now converge on the conclusion that the mind is designed to represent the prevalence of most if not all aspects of the world (objects, events, locations, etc.) in terms of natural frequencies.

Questions about the design of the human mind

There are a few important issues that are highlighted by the present research results; issues that reflect unresolved questions about how exactly the human mind processes, organizes, and stores information. One issue is the potential overloading of hypothesized natural sampling cognitive mechanisms by too much information. For example, suppose my mind tracks the frequency of females (or males) in the environment, noting those that are possible dating partners based on cues such as age, wedding rings, or other features. A memory system based on storing this information in a natural frequency structure would be useful, but could also entail huge memory demands (with hundreds of thousands of observations). These memory demands are exacerbated by the idea that my mind engages in a similar tracking of natural frequencies for any number of other aspects of the environment. Is it not necessary at some point to reign in these mental databases of natural frequencies? Several possible solutions to this dilemma may exist, including: (a) natural frequencies may become subconscious and begin to generate conscious outputs that are less precise (e.g. attitudes and feelings about situations), and (b) information that is chronologically older may be consolidated into some form of summary, much like the transfer of episodic memories to semantic memory. Whatever the case may be, this issue remains to be addressed.

A second issue highlighted by this research is the relationships between different types of statistical formats that are all nominally frequencies (e.g. simple frequencies, relative frequencies, and absolute frequencies). Relative frequencies were found to be somewhat clear and easy to understand (Experiment 1), although not as clear as simple frequencies (Experiment 4). On the one hand, it seems sensible for relative frequencies to be easy for people to understand, as they use a fairly small reference class of 100 and are often easily converted to simple frequencies in practice. Gigerenzer and Hoffrage (1995, 1999), however, consider relative frequencies to be computationally equivalent to a standard probability format because they represent normalized data. So where do relative frequencies actually fit in relation to natural frequencies and single-event probabilities? Although there are certainly several positions to take, I would suggest that although relative frequencies are, strictly speaking, computationally the same as standard probabilities, they are sometimes interpreted in practice as something like simple frequencies.⁶ In other words, it seems that many people when they see '50% off sale' may encode this information as '\$50 off every \$100 spent'. This interpretation also helps illuminate the results of Girotto and Gonzalez (2000), in which chances (e.g. 5 chances out of 100) are categorized as probabilities and relatively good statistical reasoning with such chances are taken as evidence against the frequency assumption. Such 'chances', especially when placed in a natural sampling framework (subset relations in Girotto & Gonzalez, 2000), may manage to evoke an interpretation

⁶Relative frequencies may be particularly prone to being (re)interpreted as simple frequencies in situations such as those in these studies: Situations that involve evaluating a single variable (and hence do not involve Bayesian inferences). In contrast, when two related variables (e.g. a symptom and a disease) are presented in relation to one another, relative frequencies may or may not conform to a natural sampling structure and thereupon be more or less available for translation into simple frequencies normalized to 100.

as natural frequencies. Similarly, Macchi (2000) found facilitation of statistical reasoning with materials that were a combination of simple frequencies (e.g. '360 out of every 1000 students') and percentages (e.g. 75% of the students). This finding may be due to starting the problem with simple frequencies, leading many participants to convert the subsequent percentages into a natural frequency structure based on the initial 1000-student reference class.

Finally, a third issue has to do with the issue of how very large quantities are perceived. In the introduction an analogy was drawn between the relative difficulty in discriminating between two very large quantities (e.g. 90 million and 180 million) and the scaling effects found in perception. The issue for future consideration, in essence, is to what extent is this just a useful descriptive analogy versus an indication of common underlying psychophysical processes?

Making statistics more influential

These results regarding the perception of different statistical representation formats, in addition to helping resolve a theoretical question in the field of judgments under uncertainty, lend themselves quite directly to applied uses. When the goal is to promote fast and easy understanding of the numbers, one should use small-scale (simple frequency) and percentage (relative frequency) formats. This guideline can be used in presenting information to legal juries, in designing advertisements, in presenting health and environmental risk information, in more effectively educating students, and other areas. Supplementing such information with absolute frequency (or at least absolute reference class) information should tend to produce understanding that is both clear and accurate.

There are sometimes situations, however, in which those presenting statistical information do not necessarily want to create an accurate and clear understanding of the data. For a variety of reasons, those providing the information may have prescriptive agendas that lead them to accentuate a particular perception. Under such circumstances, the results of this study indicate that these information providers may be particularly well served (or particularly ill served) by the use of absolute frequency information. Specifically, information that involves a proportionately small number of people can be made to seem more important if that proportion is expressed as an absolute frequency of a large reference class, such as the US population or the world population. On the other hand, one should avoid expressing comparative information about proportionately large numbers of people as absolute frequencies on a large scale; differences between numbers like 90 million and 180 million seem to be poorly perceived, and other presentation formats are likely to be more effective. In summary, one may or may not want to present information in an ecologically rational way, depending of the social objectives that the presentation of that information are intended to serve.

Traditionally, psychological studies of statistical information in messages designed to persuade—such as the text of advertisements—have looked at the number of information 'bits' and the empirical quality of the information. Although there are some indications that more bits of information, more positive information, and better quality information make messages more persuasive, almost no research has been done on the actual format of the statistical evidence (Reinard, 1988). In fact, the majority of past research on persuasion and attitude change has focused on the characteristics of the message recipient (e.g. personality traits, level of interest, etc.) and of the messenger (e.g. perceived expertise, attractiveness, etc.; Petty, Wegener, & Fabrigar, 1997). None of these factors deal directly with the format of statistical information that may be presented in efforts to persuade. The present results indicate, however, that principles of ecological rationality can be fruitfully applied to the field of social influence.

The general guidelines given above as to when certain statistical representation styles are and are not preferred formats in terms of persuasional impact represents a neglected aspect of the social influence literature. An increasing amount of information that people access (e.g. information from the Internet) is dissociated from many of the traditional factors associated with persuasion research (e.g. attractiveness and social approval). The objective characteristics of messages are likely to become more important as the messengers

becomes more and more removed. To some extent, one can see this phenomenon already occurring in the product of copywriters, who may use their own intuitions to select appropriate statistical formats for their purposes. Thus, for example, a recent press release from the World Health Organization (WHO) began with the following sentence:

The number of hip fractures worldwide due to osteoporosis is expected to rise threefold by the middle of the next century, from 1.7 million in 1990 to 6.3 million by 2050. (WHO, 1999)

The world held about 5.3 billion people in 1990, and the estimated world population in 2050 is 8.9 billion. If the above numbers are converted to relative percentages, the rise in the number of hip fractures is from 0.03% to 0.07% of the world population (which seems less impressive). The objective of the WHO, however, is to promote health and healthy behaviors, and it therefore makes sense to phrase these data in the most impactful way. Understanding how and why these numbers influence people is important not only for producing influential information, but also potentially important for exposing the undue influence of numbers effectively presented for less benevolent purposes.

APPENDIX: STATISTICAL INFORMATION AND PROBE QUESTIONS GIVEN TO PARTICIPANTS IN EXPERIMENT 1

It is estimated that by the year 2020, [A/B/C/D]

- A any given American will have a probability of 0.01 of having
- B 1 of every 100 Americans will have
- C 1% of all Americans will have
- D 2.7 million of all Americans will have

been exposed to Flu strain X.

1. How clear and easy to understand is the statistical information presented in the above sentence?
2. How serious do you think the existence of this virus is?
3. If you were in charge of the annual budget for the U.S. Department of Health, how much of every \$100 would you dedicate to dealing with virus X?
_____ out of every \$100

According to estimates based on a recent poll [A/B/C/D]

- A any given American has a probability of 0.33 of not knowing
- B 1 of every 3 Americans does not know
- C 33% of all Americans do not know
- D 90 million of all Americans do not know

the capital of their own state.

1. How clear and easy to understand is the statistical information presented in the above sentence?
2. How upsetting do you think this ignorance is?
3. If you were in charge of a large foundation to educate American on geographical and political issues, how much of every \$100 would you dedicate to dealing with this ignorance?
_____ out of every \$100

Based on a recent study, it was found that [A/B/C/D]

- A any given American has a probability of 0.66 of preferring
- B 2 of every 3 Americans prefer
- C 66% of all Americans prefer
- D 180 million of all Americans prefer

Soda A over Soda B.

1. How clear and easy to understand is the statistical information presented in the above sentence?
2. How good do you think Soda A is compared to Soda B? (circle one)
3. If you were in charge of a large stock portfolio, how much of every \$100 would you invest in the company producing Soda A?
_____ out of every \$100

A drug about to be approved by the FDA has been estimated as [A/B/C/D]
 A having a probability of 0.99 for being effective for any given American
 B being effective in 99 of every 100 Americans
 C being effective in 99% of all Americans
 D being effective in 263 million of all Americans
 for relieving stress.

1. How clear and easy to understand is the statistical information presented in the above sentence? (circle one)
2. How useful do you think this drug will be? (circle one)
3. If you were in charge of the production budget for the manufacturer of this drug, how much of every \$100 would you dedicate to producing this drug?
_____ out of every \$100

ACKNOWLEDGMENTS

The author would like to thank Sandra Brase and Linda Henkel for their advice and support, Ulrich Hoffrage and X. T. Wang, and an anonymous reviewer for many useful and insightful comments, and Martha Willoughby for assistance in collecting data.

REFERENCES

- Aitchison, J., & Dunsmore, I. (1975). *Statistical prediction analysis*. Cambridge: Cambridge University Press.
- Brase, G. L. (in press). 'Bugs' built into the system: an evolutionary explanation for developmental difficulties in learning about fractions. *Learning and Individual Differences*.
- Brase, G. L., Cosmides, L., & Tooby, J. (1998). Individuation, counting and statistical inference: the role of frequency and whole-object representations in judgment under uncertainty. *Journal of Experimental Psychology: General*, *127*, 3–21.
- Brown, R. D., & Newman, D. L. (1982). An investigation of the effect of different data presentation formats and order of arguments in a simulated adversary evaluation. *Educational Evaluation & Policy Analysis*, *4*, 197–203.
- Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions of the literature on judgment under uncertainty. *Cognition*, *58*, 1–73.
- Evans, J. S. B. T., Handley, S. J., Perham, N., Over, D. E., & Thompson, V. A. (2000). Frequency versus probability formats in statistical word problems. *Cognition*, *77*, 197–213.
- Fetherstonhaugh, D., Slovic, P., Johnson, S. M., & Friedrich, J. (1998). Insensitivity to the value of human life: a study of psychophysical numbing. In H. J. Langholtz (Ed.), *The psychology of peacekeeping* (pp. 75–88). Westport, CT: Praeger Publishers/Greenwood Publishing Group, Inc.
- Freudenburg, W. R., & Rursch, J. A. (1994). The risks of 'putting the numbers in context': a cautionary tale. *Risk Analysis*, *14*, 949–958.
- Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computations. *Cognition*, *44*, 43–74.
- Geary, D. C. (1995). Reflection of evolution and culture in children's cognition: implications for mathematical development and instruction. *American Psychologist*, *50*, 24–37.

- Gelman, R. (1998). Domain specificity in cognitive development: universals and nonuniversals. In M. Sabourin, & F. Craik (Eds.), *Advances in psychological science, Vol. 2: Biological and cognitive aspects* (pp. 557–579). Hove: Psychology Press/Erlbaum.
- Gigerenzer, G. (1991). How to make cognitive illusions disappear: beyond heuristics and biases. *European Review of Social Psychology, 2*, 83–115.
- Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: frequency formats. *Psychological Review, 102*, 684–704.
- Gigerenzer, G., & Hoffrage, U. (1999). Overcoming difficulties in Bayesian reasoning: a reply to Lewis and Keren (1999) and Mellers and McGraw (1999). *Psychological Review, 106*, 425–430.
- Gigerenzer, G., Hoffrage, U., & Ebert, A. (1998). AIDS counseling for low-risk clients. *AIDS Care, 10*, 197–211.
- Gigerenzer, G., Todd, P. M., & The ABC Research Group. (1999). *Simple heuristics that make us smart*. Oxford: Oxford University Press.
- Giroto, V., & Gonzalez, M. (2001). Solving probabilistic and statistical problems: a matter of information structure and question form. *Cognition, 78*, 247–276.
- Halpern, D. F., Blackman, S., & Salzman, B. (1989). Using statistical risk information to assess oral contraceptive safety. *Applied Cognitive Psychology, 3*, 251–260.
- Harries, G., Harvey, N. (2000). Are absolute frequencies, relative frequencies, or both effective in reducing cognitive biases? *Journal of Behavioral Decision Making, 13*, 431–444.
- Hartnett, P., & Gelman, R. (1998). Early understandings of numbers: paths or barriers to the construction of new understandings? *Learning & Instruction, 8*, 341–374.
- Hoffrage, U., Gigerenzer, G., Krauss, S., & Martingon, L. (in press). Representation facilitates reasoning: what natural frequencies are and what they are not. *Cognition*.
- Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Communicating statistical information. *Science, 290*, 2261–2262.
- Howson, C., & Urbach, P. (1993). *Scientific reasoning: The Bayesian approach* (2nd edn). Chicago: La Salle.
- Johnson-Laird, P. N., Legrenzi, P., Giroto, V., Legrenzi, M. S., & Caverni, J. P. (1999). Naive probability: a mental model theory of extensional reasoning. *Psychological Review, 106*, 62–88.
- Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, MA: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1972). Subjective probability: a judgment of representativeness. *Cognitive Psychology, 3*, 430–454.
- Kleiter, G. (1994). Natural sampling: rationality without base rates. In G. H. Fischer, & D. Laming (Eds.), *Contributions to mathematical psychology, psychometrics, and methodology* (pp. 375–388). New York: Springer-Verlag.
- Macchi, L. (2000). Partitive formulation of information in probabilistic problems: beyond heuristics and frequency format explanations. *Organizational Behavior and Human Decision Processes, 82*, 217–236.
- Matlin, M. W. (1994). *Cognition* (4th edn). Fort Worth, TX: Harcourt Brace College Publishers.
- National Public Radio. 13 June 2000. Wage gap. *The Diane Rehm Show*.
- Petty, R. E., Wegener, D. T., & Fabrigar, L. R. (1997). Attitudes and attitude change. *Annual Review of Psychology, 48*, 609–647.
- Reinard, J. C. (1988). The empirical study of the persuasive effects of evidence: the status after fifty years of research. *Human Communication Research, 15*, 3–59.
- Samuels, R., Stich, S., & Bishop, M. (2001). Ending the rationality wars: how to make disputes about human rationality disappear. In R. Elie (Ed.), *Common sense, reasoning and rationality, Vancouver Studies in Cognitive Science, Vol. 11*. New York: Oxford University Press.
- Shepard, R. N. (1984). Ecological constraints on internal representation: resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychological Review, 91*, 417–447.
- Shepard, R. N. (1987). Evolution of a mesh between principles of the mind and regularities of the world. In J. Dupre (Ed.), *The latest on the best: Essays on evolution and optimality* (pp. 251–275). Cambridge, MA: The MIT Press.
- Simon, H. A. (1956). Rational choice and the structure of environments. *Psychological Review, 63*, 129–138.
- Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology, 41*, 1–19.
- Thompson, W. C., & Schumann, E. L. (1987). Interpretation of statistical evidence in criminal trials: the prosecutor's fallacy and the defense attorney's fallacy. *Law & Human Behavior, 11*, 167–187.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*, 453–458.
- Wang, X. T. (1996a). Domain-specific rationality in human choices: violations of utility axioms and social contexts. *Cognition, 60*, 31–63.
- Wang, X. T. (1996b). Evolutionary hypotheses of risk-sensitive choice: age differences and perspective change. *Ethology & Sociobiology, 17*, 1–15.

- Wang, X. T. (1996c). Framing effects: dynamics and task domains. *Organizational Behavior & Human Decision Processes*, 68, 145–157.
- Wang, X. T., & Johnston, V. S. (1995). Perceived social context and risk preference: a re-examination of framing effects in a life–death decision problem. *Journal of Behavioral Decision Making*, 8, 279–293.
- World Health Organization Press Release WHO/58: Osteoporosis: both health organizations and individuals must act now to avoid an impending epidemic. (1999). <http://www.who.org/inf-pr-1999/en/pr99-58.html> [11 October 1999].
- Wynn, K. (1992). Issues concerning a nativist theory of numerical knowledge. *Mind & Language*, 7, 367–381.
- Wynn, K. (1995). Infants possess a system of numerical knowledge. *Current Directions in Psychological Science*, 4, 172–177.
- Wynn, K. (1998a). An evolved capacity for number. In D. D. Cummins, & C. Allen (Eds.), *The evolution of mind* (pp. 107–126). New York: Oxford University Press.
- Wynn, K. (1998b). Numerical competence in infants. In C. Donlan (Ed.), *The development of mathematical skills: Studies in developmental psychology* (pp. 3–25). Hove: Psychology Press.

Author's biography:

Gary L. Brase is currently a Senior Lecturer in Psychology at the University of Sunderland. His research covers adaptive reasoning and decision making within statistical and social domains.

Author's address:

Gary L. Brase, Sunderland Business School, Division of Psychology, University of Sunderland, St Peter's Campus, St Peter's Way, Sunderland, SR6 0DD, UK.