Journal of Experimental Psychology: General
2002, Vol. 131, No. 3, 424–442

0096-3445/02/$5.00 DOI: 10.1037//0096-3445.131.3.424

What Very Small Numbers Mean

Dale J. Cohen, Jennifer M. Ferrell, and Nathan Johnson
University of North Carolina at Wilmington

This article presents a theoretical and experimental framework for assessing the biases associated with the interpretation of numbers. This framework consists of having participants convert between different representations of quantities. These representations should include both variations in numerical labels that symbolize quantities and variations in displays in which quantity is inherent. Five experiments assessed how people convert between relative frequencies, decimals, and displays of dots that denote very low proportions (i.e., proportions below 1%). The participants demonstrated perceptual, response, and numerical transformation biases. Furthermore, the data suggest that relative frequencies and decimals are associated with different abstract representations of amount.

Scientists and lay people use several *numerical formats* (i.e., any symbol system used to represent quantities) to symbolize proportions. For example, the decimal "0.5" and relative frequency "1 in 2" symbolize the same proportion. Although relative frequencies and decimals denote proportions equally well, people may interpret these numerical formats differently. Nevertheless, researchers often make an implicit assumption that people interpret these numerical formats equivalently (termed the *assumption of numerical equivalence*). The implicit assumption of numerical equivalence is prevalent in studies of psychophysics (e.g., Gescheider, 1988; Marks, 1974; Marks & Algom, 1998; Stevens, 1956, 1986), risk estimation (e.g., D. J. Cohen & Bruce, 1997; Gladis, Michela, Walter, & Vaughan, 1992; Hansen, Hahn, & Wolkenstein, 1990; Mickler, 1993; van der Velde, van der Plight, & Hooykaas, 1994), mathematics (e.g., Ashcroft, 1992; Ashcroft & Kirk, 2001; Campbell & Xue, 2001), reasoning (e.g., Kahneman & Tversky, 1972; Tversky & Kahneman, 1974; Wanke, Schwarz, & Bless, 1995), and probability estimation (e.g., Begg, 1974; Brooke & MacRae, 1977; Hollands & Dyre, 2000; Shuford, 1961; Spence, 1990; Teigen, 1973; Tversky & Fox, 1995; Varey, Mellers, & Birnbaum, 1990), to name just a few. The assumption of numerical equivalence, however, should not be made lightly because people's interpretation of numbers likely affects the data they produce, and thus the conclusions that researchers draw.

In this article, we present a theoretical and experimental framework for assessing the biases associated with how people interpret numbers, and we show how this framework can reveal details concerning how quantity information is represented in the human brain. We report five experiments that assess how people interpret and represent relative frequencies and decimals that denote very low proportions (i.e., proportions below 1%). We chose to examine very low proportions for three reasons: (a) There is extensive evidence that people do not interpret *numerical labels* (i.e., any individual symbol representing quantity) symbolizing very low proportions accurately (for a review, see Rothman & Kiviniemi, 1999), (b) numerical labels symbolizing very low proportions are vital to the communication of issues ranging from health risks to pollution (e.g., Environment Protection Agency, 1991), and (c) people's interpretation of relatively common quantities (between 1% and 99%) has been studied (e.g., see Begg, 1974; Brooke & MacRae, 1977; Hollands & Dyre, 2000; Shuford, 1961; Spence, 1990; Teigen, 1973; Tversky & Fox, 1995; Varey et al., 1990) and therefore can serve as a comparison group.

Estimating Proportions

Most psychophysical research on the relation between perceived and estimated proportions has concentrated on quantities between 1% and 99% (e.g., see Begg, 1974; Brooke & MacRae, 1977; Shuford, 1961; Teigen, 1973; Tversky & Fox, 1995; Varey et al., 1990). Generally, researchers present different auditory or visual stimuli and ask participants to estimate the proportion of a target stimulus. Most researchers have found an inverse ogival relation (inverted S-shaped curve) between the reported and actual proportions, in which participants overestimate proportions below 50% and underestimate proportions above 50% (for a review, see Hollands & Dyre, 2000).

A frequently used procedure for investigating proportion estimation is the *judgment task*. In the judgment task, participants are presented with two stimuli and are asked to judge the proportion of one stimulus in relation to the total set of stimuli. Varey et al. (1990) provided a typical example. Varey et al. created 36 different displays of randomly placed black and white dots, in which the proportion of white dots ranged between 1.6% and 50%. Varey et al. varied the absolute number of black and white dots in the displays so the absolute and relative frequencies of the target were not confounded. Participants estimated the percentage of white dots, the percentage of black dots, the difference between black dots and white dots, and the ratio of black dots to white dots. The data revealed the inverse ogival error pattern found in the earlier studies, demonstrating the robustness of this pattern in proportion estimation.

Dale J. Cohen, Jennifer M. Ferrell, and Nathan Johnson, Department of Psychology, University of North Carolina at Wilmington.

This article was partially supported by a Charles Cahill grant from the University of North Carolina at Wilmington. We thank Steffaney Cohen, Nikole Eagen, Julian Keith, Len Lecci, and William Overman for their helpful comments and support at various stages of this project.

Correspondence concerning this article should be addressed to Dale J. Cohen, Department of Psychology, University of North Carolina at Wilmington, Wilmington, North Carolina 28403-3297. E-mail: cohend@uncwil.edu

Although the judgment task can be used to demonstrate the pattern of estimation error, alone it cannot reveal the source of that error. In an earlier study, Brooke and MacRae (1977) attempted to determine the source of the inverse ogival error pattern. In the first part of their study, participants were presented displays containing 100 horizontal and vertical lines and were asked to estimate the proportion of horizontal to total lines (i.e., a judgment task). In the second part of their study, participants were shown numerical proportions and were asked to produce displays of horizontal and vertical lines that represented those proportions (termed a *production* task). The proportions used in both experiments ranged from 10% to 90%, in multiples of 10%. The researchers explained that if the inverse ogival error pattern were due to a response bias, the patterns for the judgment and production tasks would be similar. If the inverse ogival error pattern were due to a perceptual bias, the patterns for the judgment and production tasks would be inversely related (i.e., inverse ogival vs. ogival). The data revealed inversely related patterns, leading Brooke and MacRae to conclude that a perceptual bias was the source of the inverse ogival error pattern.

Typically, perceptual biases follow Stevens's Power Law (Stevens, 1986). Stevens's Power Law states that the perceptual magnitude of a stimulus, Ψ, is a function of physical magnitude, ϕ, taken to a power, β:

$$\Psi = k\phi^{\beta}, \tag{1}$$

where k is a function of the units of measurement and β "serves as a signature that may differ from one sensory continuum to another" (Stevens, 1986, p. 13). Spence (1990) showed that a variant of Stevens's Power Law can model the inverse ogival error pattern generally found in proportion judgment tasks. Spence proposed that participants estimate the magnitude of both proportions in the display (P and Q) and that participants constrain their estimates to ensure these quantities sum to 1. On the basis of this supposition, Spence derived a model of proportion estimation (termed the *power model* by Hollands & Dyre, 2000), which states that the perceptual magnitude of a proportion, Ψ, is a function of the proportion, ϕ, and its inverse, $1- \phi$:

$$\Psi = \phi^{\beta}/[(\phi^{\beta}) + (1 - \phi)^{\beta}]. \tag{2}$$

Spence (1990) showed that the power model successfully describes the inverse ogival error pattern present in participants' estimates of a variety of stimuli.

Hollands and Dyre (2000) generalized the power model to accommodate data with multiple ogival cycles. Hollands and Dyre proposed that people judge proportions in comparison with reference points. The single-cycle inverse ogival pattern results when people use 0 and 1 as reference points (Spence, 1990). Hollands and Dyre suggested that people can and do adopt reference points other than 0 and 1. Hollands and Dyre formalized this theory in their *cyclical power model*.

Although several studies have demonstrated the robustness of the inverse ogival error pattern in proportion estimation, these experiments have only examined proportions between 1% and 99%. It is unclear whether the biases associated with estimating these values would generalize to more extreme values (e.g., less than .01).

On Physical Stimuli and Numerical Responses

The articles discussed earlier focused on perceptual mechanisms (i.e., errors in perceiving the stimulus) as the source of the bias in proportion judgment. Although there is undoubtedly a perceptual component, there is also likely a cognitive component (i.e., errors in associating a number with an amount and vice versa). Shepard (1981) addressed this issue in his discussion of the information inherent in participants' estimates of magnitude. Shepard stated that the transformation from stimulus to response occurs in two stages. In the first stage, the physical stimulus is transformed into a psychological experience of the stimulus, termed the *stimulus transformation function*:

$$\Psi = f_S(\phi). \tag{3}$$

In the second stage, the psychological experience of the stimulus is transformed into a response, termed the *response transformation function*. When the response is a number (#), the response transformation function may more specifically be termed a *numerical transformation function*:

$$\# = f_N(\Psi). \tag{4}$$

Because researchers only have access to a participant's response, it is not possible to assess either of the two functions in isolation. Therefore, the function describing the relation between a participant's numerical response and the physical stimulus is the combined stimulus and numerical transformation functions:

$$\# = g(\phi), \tag{5}$$

where $g = f_N f_S$.

Although one cannot assess either function in isolation, one can extract some information about each function by simultaneously fixing one function (e.g., f_S) and manipulating the other function (e.g., f_N). For example, one can present two groups of participants the same stimulus and ask each group to respond with a different numerical format. Because the same stimulus is presented to all participants, the stimulus transformation function is fixed. Therefore, if participants' responses differ as a function of numerical format, one can conclude that the difference is the result of different numerical transformation functions. The greater the effect of numerical format, the greater the impact of the numerical transformation functions.

Although it is possible to experimentally assess the impact of f_N and f_S, psychology researchers have tended to assume that there is little or no effect of f_N (i.e., $f_N = 1$). Recent data, however, cast doubt on that assumption. Sedlmeier and Gigerenzer (2001) proposed that the degree to which people understand numerical labels depends on the type of label used. They argued that humans are "tuned" to process amount information expressed as relative frequencies (e.g., 1 in *X*) in the same way that calculators are tuned to process amount information expressed in Arabic base 10 format. When amount information is expressed in a format other than relative frequencies (e.g., decimal format), humans will fail to process the amount information accurately. Sedlmeier and Gigerenzer (2001) and Gigerenzer and Hoffrage (1995) supported their claims by demonstrating that participants engage in accurate Bayesian reasoning when amount information is expressed as relative frequencies but they fail at Bayesian reasoning when amount information is expressed in decimal format. Prior research assessed Bayesian reasoning using decimal format and, based on the implicit assumption that $f_N = 1$, concluded that humans were inherently poor Bayesian reasoners (for a review, see Sedlmeier & Gigerenzer, 2001). Gigerenzer and his colleagues have shown that effective Bayesian reasoning depends on f_N, not inherent reasoning abilities.

On Numerical Formats and Quantity Representations

Understanding the relation between f_N and different numerical formats is also central to the assessment of how numbers are represented in the human brain. Specifically, the variability of f_N as a function numerical format likely arises because amount information is not inherent in numerical labels. Numerical labels are an element of language, therefore any relation between these labels and amount information must be indirect. This suggests that there exists a language-independent abstract representation of amount (termed a *quantity representation*).

Recent neuroimaging research supports the existence of a quantity representation. Using a functional magnetic resonance imaging technique, Dehaene, Spelke, Pinel, Stanescu, and Tsivkin (1999) showed that quantity representations are processed in visuospatial networks of the left and right parietal lobes, whereas numerical labels are processed in the "left inferior frontal circuit also used for generating associations between words" (p. 973). That is, numerical labels are processed in areas of the brain connected with language, and activation of these areas by numerical labels does not automatically activate the visuospatial areas of the brain that represent quantity. These findings have been supported in tests using positron emission tomography (Dehaene et al., 1996) and by examining brain-damaged patients (L. Cohen, Dehaene, Chochon, Lehericy, & Naccache, 2000).

The identification of a brain area associated with a quantity representation may lead one to conclude that a single representational format exists to which all numerical labels, regardless of format, link. McCloskey, Caramazza, and Basili (1985) accepted this hypothesis and proposed that this single quantity representation mediates all numerical comprehension, production, and computational processes (see also McCloskey, 1992). That is, people do not understand, produce, or manipulate numbers without activating this quantity representation.

Support for a single quantity representation is provided by experiments that assess whether different numerical formats show the same pattern of results in number comparison tasks. For example, when participants are asked to determine which of two simultaneously presented numbers is numerically larger, there is a monotonically decreasing function relating reaction time (RT) and the numerical distance between the two numbers (termed the *numerical distance effect*; Moyer & Landauer, 1967). The numerical distance effect has been found with verbal (e.g., "two"; Dehaene & Akhavein, 1995; W. Schwarz & Ischebeck, 2000), Arabic (e.g., "2"; Dehaene & Akhavein, 1995; W. Schwartz & Ischebeck, 2000), and Japanese kanji and kana numerals (Takahashi & Green, 1983). Furthermore, Buckley and Gillman (1974) provided evidence that the same process underlies the comparison of Arabic numerical labels and the comparison of dots. Finally, W. Schwarz and Ischebeck (2000) demonstrated that sequential priming of early visual processing, the lexicon, and the phonological representation resulted in predictable reductions in RT. The similarity of results across numerical formats provides indirect evidence that verbal, Arabic, Japanese kanji and kana numerals, and displays of dots share a common quantity representation.

In contrast to McCloskey's (1992) model, Gonzalez and Kolers (1982) proposed that different numerical formats are associated with different quantity representations. Gonzalez and Kolers presented an experiment in which they asked participants to verify the sum of two numbers ($a + b = c$). Each of the three numbers in the equation was either an Arabic or Roman numeral. They examined how the number of Roman numerals in the equation related to the function relating RT to the minimum addend (and the sum squared). If Arabic and Roman numerals share a representation, but participants find it more difficult to recover the representation when prompted with a Roman numeral, then the number of Roman numerals will only affect the intercept of the function. If Arabic and Roman numerals are associated with different representations, then it is likely (though not guaranteed) that the number of Roman numerals will affect the slope of the function. The results showed that as the number of Roman numerals in the equation increased, the slope relating RT to the minimum addend (and the sum squared) increased, suggesting that Arabic and Roman numerals are associated with different quantity representations (but see Sokol, McCloskey, Cohen, & Aliminosa, 1991). Gonzalez and Kolers (1982) stated that the two numerical formats may not share a common quantity representation because the different numerical formats restrict the way in which amount information can be represented. Gonzalez and Kolers further suggested that the quantity representation associated with one numerical format may be inaccessible to that associated with another numerical format.

The relation between f_N and different numerical formats can address the validity of the McCloskey model and the Gonzalez and Kolers model. Whereas both the McCloskey model and the Gonzalez and Kolers model permit f_N to vary as a function numerical format, the two models differ in their predictions concerning the conversion from one numerical format into another. The McCloskey model posits that when one converts from Format A to Format B, one first associates Format A with a quantity representation and then associates this same quantity representation with Format B. If we use Shepard's (1981) notation, then the conversion from Format A to a quantity representation would be the inverse numerical transformation function for Format A,

$$\Psi = f_{NA}^{-1}(\#_A). \tag{6}$$

The conversion from that quantity representation to Format B would be the numerical transformation function for Format B,

$$\#_B = f_{NB}(\Psi). \tag{7}$$

Therefore the bias associated with the conversion between the two numerical formats should be a combination of the individual biases,

$$\#_B = f_{NB}(f_{NA}^{-1}(\#_A)). \tag{8}$$

This hypothesis cannot be assessed with a point prediction because one does not have access to the numerical transformation functions in isolation. However, if one measures $\#_B = g(\phi)$ (i.e., the judgment task using Format B) and $\phi = g^{-1}(\#_A)$ (i.e., the production task using Format A), one can assess whether a participant's conversion data can be reasonably fit within the constraints of Equation 8.

In contrast to the McCloskey model, the Gonzalez and Kolers model posits that each numerical format is associated with a different quantity representation, and these quantity representations may be inaccessible to one and other. If the two quantity representations are inaccessible to one and other, a nonanalog process (e.g., rule-based) must be used to convert between the numerical formats.[1] Because no quantity representation is involved in the conversion process, (a) the bias associated with converting one numerical format into another cannot be predicted from the biases associated with each numerical format, and (b) participants may not notice if the nonanalog process that they use to convert between the two numerical formats is outrageously incorrect. Therefore, the McCloskey model and the Gonzalez and Kolers model make qualitatively different predictions concerning the conversion between two numerical formats.

In the present article, we assess the perceptual and cognitive biases associated with estimating very low proportions using relative frequencies and decimals. Furthermore, we use the estimates of these biases to assess whether comprehension and production of very low proportions are mediated by a single quantity representation. In Experiment 1, we presented participants with proportions represented by displays of dots and asked them to estimate those proportions in either relative frequency or decimal format (i.e., a judgment task). Experiment 1 provides an estimate of the combined stimulus and numerical transformation function, $g(\phi)$, for both decimals and relative frequencies. Because participants in the decimal and relative frequency conditions view the same stimuli, the stimulus transformation function should be fixed. Therefore, if $g(\phi)$ differs by numerical format, the difference can be ascribed to different numerical transformation functions. In Experiment 2, we presented participants with proportions in relative frequency or decimal format and asked them to create displays of dots that represent the presented proportions (i.e., a production task). Experiment 2 provides an estimate of the inverse combined stimulus and numerical transformation function, $g^{-1}(\#)$, for both decimals and relative frequencies. Again, if $g^{-1}(\#)$ differs by numerical format, the difference can be ascribed to different numerical transformation functions. In Experiments 3 and 4, we varied the total number of dots in the judgment and production tasks. Experiments 3 and 4 assessed whether $g(\phi)$ and $g^{-1}(\#)$ change as the number of dots in the display changes. If there is an effect of number of dots within each numerical format condition, the effect can be ascribed to different stimulus transformation functions. Finally, in Experiment 5, we asked participants to convert from relative frequencies to decimals and vice versa. If the participants' conversion strategies are analog based and fit within the constraints of Equation 8, the data will provide confirming evidence for the McCloskey model. If the participants' conversion strategies are nonanalog and incorrect, the data will provide confirming evidence for the Gonzalez and Kolers model. To efficiently describe the estimation process, we discuss the individual and combined results of these five experiments in the General Discussion.

Experiment 1

Experiment 1 investigated how people estimate proportions between .0001 and .01. We assessed $g(\phi)$ for decimals and relative frequencies by presenting participants displays of black and white dots and asking them to label the proportion of white dots to total dots using either decimals or relative frequencies (depending on condition). Because of the relatively robust evidence that gender-specific effects are found in spatial and mathematical reasoning tasks (for a review, see Hyde & McKinley, 1997), we included gender as a variable.

Method

Participants. Data were collected from 80 introductory psychology students enrolled at a midsized university in the southeastern United States. Volunteers were unaware of the purpose of the experiment and received class credit for participation. Forty participants were women and 40 participants were men.

Apparatus. All stimuli were presented on a 13-in. (33-cm) VGA color monitor with a 60-Hz refresh rate controlled by an 80486 microcomputer using the DOS operating system. The resolution of the monitor was 1024 × 768.

Stimuli. The term *display* used throughout the article refers to the stimuli of black and white dots against a uniform background. Each display filled a 13-in. (33-cm) computer monitor screen with a visual angle of 24.6°. Each display consisted of 50,000 dots scattered randomly on the computer screen. The dots were white or black on a red background. These colors were used because the contrast enabled participants to easily see each element on the screen. The proportion of white to total dots in each display was randomly drawn from a uniform distribution of proportions ranging from .0001 to .01.

A response screen was presented to participants after each display. The response screen instructed participants to type in their estimates of the proportion of white to total dots in the previous display. Participants in the decimal condition were asked to respond in decimals, and participants in the relative frequency condition were asked to respond in relative frequencies (i.e., 1 in *X*). The participant's response was presented in the center of the screen, and the numbers subtended a vertical visual angle of 1.33°. There was a 10-s delay between the end of the previous trial and the presentation of the next display (because of the computer generating the display for the next trial).

Procedure. Participants were randomly assigned to one of two response type conditions: decimals or relative frequency. Participants were tested individually.

Each trial consisted of the presentation of a display, the participants' estimation of the proportion of white dots to total dots in the displays, and a chance for the participants to check their response. Participants viewed the display of dots for 3 s, with the black dots appearing on the screen for 1 s before white dots. The black dots appeared first to allow participants to visually focus (i.e., accommodate) on the dots on the screen. The white and black dots were then presented together for 2 s. This 2-s display enabled the participants to clearly perceive the dots but prohibited them from counting the dots. Participants were not told the total number of dots, and a new random display was generated for each trial to prohibit participants from recognizing patterns.

After the display was presented, the response screen appeared. The response screen instructed participants to estimate the proportion of white dots to total dots in the previous display. Participants used the keyboard to respond either in relative frequency or decimal format, depending on their assigned condition. For participants in the relative frequency condition, the words "1 in" were provided, and participants filled in the rest of the relative frequency with a whole number. For example, if participants estimated the proportion to be 1 in 100, participants would type in "100." For participants in the decimal condition, the participants typed a decimal between 0 and 1. After participants responded, they were shown the number they had entered and were asked if this proportion was the number they wanted to use as their estimate of the proportion of white dots. Participants pressed the "y" key to indicate that they had entered their estimate correctly or the "n" key

[1] Nonanalog strategies have been discussed in the numerical representation literature (e.g., Campbell, 1994; Dehaene & Akhavein, 1995; Dehaene et al., 1996). This literature addresses the question of whether people can use numerical labels without accessing a quantity representation. Examples of such strategies are using memorized multiplication tables, using mathematical "tricks," and converting between numerical formats through memorized associations rather than through quantity representations.

to indicate that they had not entered their estimate correctly. If participants pressed the “n” key, they were asked for their estimate again. If participants pressed the “y” key, they proceeded to the next trial.

The sessions consisted of 10 practice trials and 150 experimental trials. After every 32 experimental trials, participants were given a self-timed rest period. Each session lasted approximately 50 min. The computer recorded participants’ estimates.

Results

When plotted with the presented proportions on the *x*-axis and estimates on the *y*-axis, each participant’s data evidenced a distinct curve. Exploratory analysis revealed that a log-log transformation best linearized the data (i.e., a power function). We therefore fit a regression line to each participants’ log-log transformed data (for an example of a typical participant in each condition, see Figures 1 and 2). To assess whether the power function was the most appropriate transformation, we compared the amount of variance accounted for (i.e., r^2) by these power functions with the amount of variance accounted for by three reasonable alternatives: (a) the untransformed data, (b) Spence’s (1990) power model, and (c) Hollands and Dyre’s (2000) cyclical power model. When computing Hollands and Drye’s (2000) cyclical power model, we set the lower-bound reference value to 0 and allowed the upper-bound reference value to exist between 0 and 1.[2] Finally, we computed a one-way within-subjects analysis of variance (ANOVA) comparing the r^2 values of each model. There was a significant effect of model, $F(3,\ 316) = 27.07$, $p < .001$, $MSE = 0.055$. Tukey’s honestly significance difference ($p < .05$) revealed that the regression lines fit to the log-log transformed data (i.e., power function) accounted for significantly more variance ($M_{r^2} = 0.56$, $SD_{r^2} = 0.22$) than the other three models. The amount of variance accounted for by the cyclical power model ($M_{r^2} = 0.34$, $SD_{r^2} = 0.27$) and that of the regression lines fit to the untransformed data ($M_{r^2} = 0.38$, $SD_{r^2} = 0.26$) did not differ significantly but were significantly higher than the amount of variance accounted for by the power model ($M_{r^2} = 0.23$, $SD_{r^2} = 0.18$). Because the power functions best fit the data, all further analyses were performed on the log-log transformed data.

We used the slopes and intercepts of each participant’s function to examine the degree of over- or underestimation of the presented proportions. The slope of the log-log transformed data determines the curvature of the function describing the raw data. A slope greater than 1 describes an accelerating function (i.e., line that curves upward), and a slope less than 1 describes a decelerating function (i.e., a line that curves downward). The intercept has been traditionally discarded in magnitude estimation experiments be-

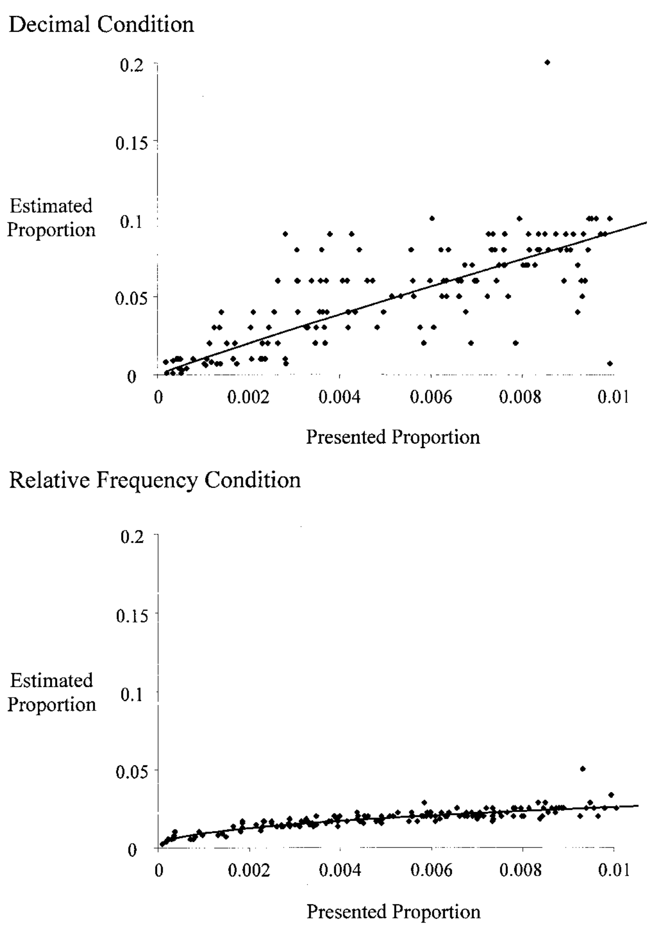

Figure 1. Plots of presented proportion by estimated proportion for a typical participant in the decimal and relative frequency conditions of Experiment 1. Power functions are fit to the data.

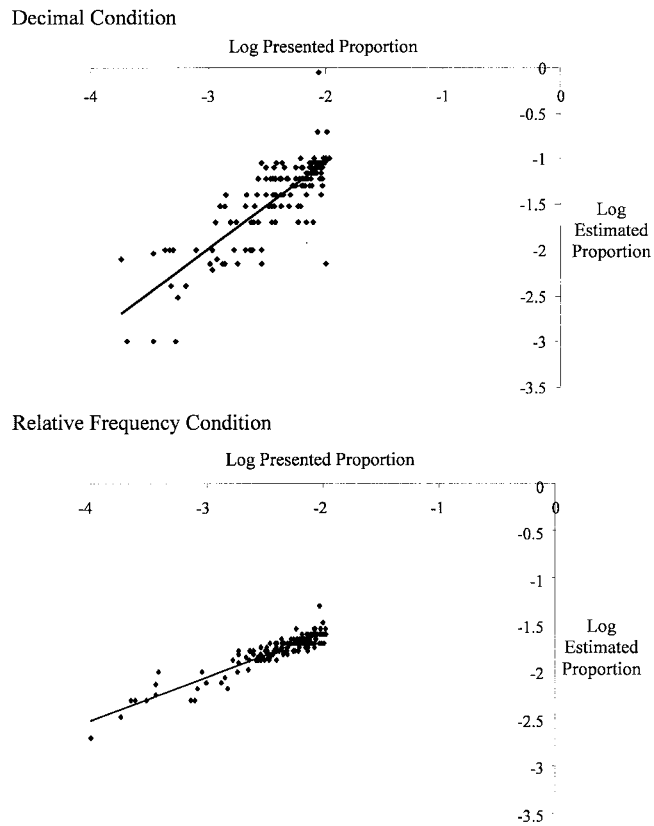

Figure 2. Log-log plots of presented proportion by estimated proportion for the participants presented in Figure 1. Linear functions are fit to the data.

[2] We fit both the power model and the cyclical power model using the multivariate secant iterative method (SAS Institute, 1994).

cause it is a function of the value of the standard (Stevens, 1986). However, in a proportion estimation experiment, the intercept describes, in part, the magnitude of the response. For example, if the intercept of two otherwise identical functions differs by 1 point, then the predictions of the two functions differ by one order of magnitude. Therefore, we also included the intercept as a dependent variable. Finally, the R^2 value of each participant was used as a measure of the consistency of the pattern of each participant's estimates.

Means and standard deviations for the slopes, intercepts, and R^2 values by response type and gender for Experiment 1 are presented in Table 1. A 2 × 2 between-subjects ANOVA for response type (relative frequency vs. decimal) by gender was performed on each of the three dependent variables (i.e., slopes, intercepts, and R^2 values).

There was a significant main effect of response type on slopes, $F(1, 76) = 19.44$, $p < .001$, $MSE = 0.31$, in which participants who estimated proportions as decimals had a higher slope than those estimating proportions in relative frequencies. Distributions of participants' slopes for each response type are presented in Figure 3. There were no other significant effects on slopes.

There was a significant main effect of response type on intercepts, $F(1, 76) = 35.90$, $p < .001$, $MSE = 1.67$, in which participants who estimated proportions as decimals had a higher intercept than those estimating proportions in relative frequencies. Distributions of participants' intercepts for each response type are presented in Figure 3. There were no other significant effects on intercepts.

There was no significant main effect of response type on R^2 values, $F(1, 76) = 3.62$, $p = .06$, $MSE = 0.04$. There was a significant main effect of gender on R^2 values, $F(1, 76) = 7.92$, $p = .006$, $MSE = 0.04$, in which female participants' R^2 values were higher than male participants' R^2 values. There was a significant interaction between gender and response type for the R^2 values, $F(1, 76) = 8.58$, $p = .005$, $MSE = 0.04$, such that in the relative frequency condition, male and female participants did not differ, but in the decimal condition, female participants had higher R^2 values than did male participants. The participants' average functions for each response type can be seen in Figure 4.

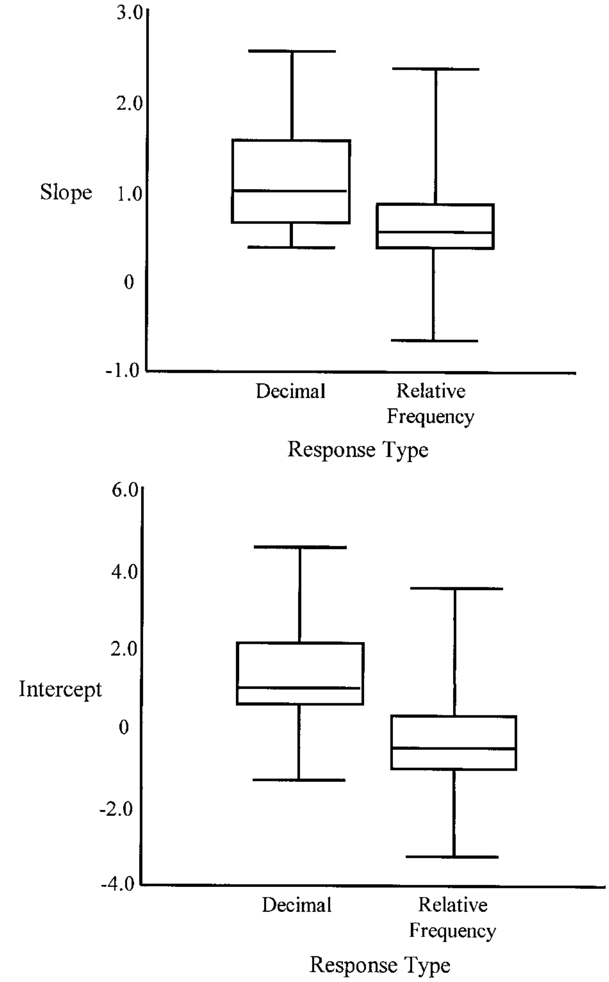

Figure 3. Box plots of the slopes and intercepts in the decimal and relative frequency conditions of Experiment 1.

Discussion

The purpose of Experiment 1 was to examine the relation between presented proportions between .0001 and .01 and participants' estimates of those proportions. In addition, we investigated whether gender or response type affected this relation. There were four important findings. First, in both the relative frequency and decimal conditions, participants overestimated the very low proportions in the displays. Second, the relation between the presented proportions and participants' estimates was shown to be a power function. Third, although participants using the two response types viewed the same stimuli, each response type was associated with a different pattern of error. Finally, there was no effect of gender. We discuss the possible causes and implications of these patterns in the General Discussion.[3]

Table 1
Means and Standard Deviations of Slopes, Intercepts, and R^2 Values in Experiment 1 by Response Type and Gender

Gender and response type	Slope		Intercept		R^2	
	M	*SD*	*M*	*SD*	*M*	*SD*
Male						
Decimal	1.231	0.639	0.917	1.490	0.477	0.265
Relative frequency	0.634	0.280	−0.421	0.963	0.523	0.237
Female						
Decimal	1.116	0.650	1.728	1.204	0.731	0.080
Relative frequency	0.616	0.574	−0.396	1.442	0.518	0.156

Experiment 2

Experiment 2 was designed to assesses $g^{-1}(\#)$. Whereas in Experiment 1 participants were asked to label the proportion of white dots to total dots in a display, in Experiment 2 participants were given a proportion expressed as a decimal or relative frequency and were asked to produce a display in which the proportion of white dots to total dots matched the presented proportion. If participants' estimates are only affected by stimulus and numerical transformation functions,

[3] Because the R^2 results indicated that all of the participants' responses were well described by the regressions, the interaction between gender and response type for the R^2 values has no important impact on the interpretation of participants' estimates. We therefore do not discuss them further.

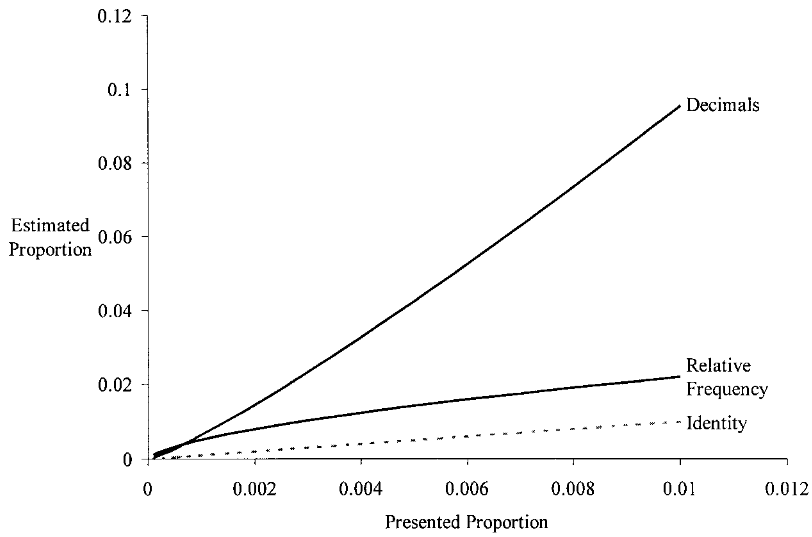

Figure 4. The best fit functions of estimated proportions, or *f*(presented proportions), for participants in the decimal and relative frequency conditions of Experiment 1. These functions were calculated using the average slope and intercept across participants and are plotted on raw scales so the curvature of the functions can be seen. Also included is the identity line (i.e., slope = 1.0).

then the results of Experiments 1 and 2 should be inversely related. If the results of Experiments 1 and 2 are not inversely related, then it is likely that participants' estimates are influenced by a response bias (see Brooke & MacRae, 1977). Finally, we examined gender differences in Experiment 2, although no gender differences were found in Experiment 1.

In Experiment 2 we also included a condition in which participants simply produce a display to match the presented display. The display condition in the present experiment can be characterized by sequential application of the stimulus transformation function, $\Psi = f_S(\phi)$, and the inverse stimulus transformation function, $\phi = f_S^{-1}(\Psi)$. That is,

$$\phi = f_S^{-1}(f_S(\phi)). \tag{9}$$

Therefore, any deviation from accuracy can be attributed to other unidentified effects, such as memory.

Method

Participants. One hundred and twelve participants were drawn from the same population as in Experiment 1. Volunteers received class credit for participation. Sixty-five participants were women and 47 participants were men.

Stimuli. The rooms and computers used in this experiment were the same as those used in Experiment 1. On each trial, participants were presented with a proportion expressed as a decimal, a relative frequency, or a display. The proportions were randomly drawn from a uniform distribution ranging from .0001 to .01. The decimal and relative frequency proportions were centered on the computer screen, and the numbers subtended a vertical visual angle of 1.33°. The displays in the display condition were identical to those used in Experiment 1.

After the proportion was removed, participants were presented a response screen consisting of 50,000 black dots. These displays were generated in the same way as in Experiment 1, with the exception that they contained no white dots. Participants adjusted the proportion of white dots in the display by pressing number keys on the keyboard of the computer. The "3" key added 1 white dot, the "6" key added 10 white dots, and the "9" key added 100 white dots. The "1" key removed 1 white dot, the "4" key removed 10 white dots, and the "7" key removed 100 white dots. Each white dot added to the display replaced a random black dot in the display (and vice versa) so that the total number of dots remained the same in each trial.

Procedure. Participants were randomly assigned to one of three presentation types: decimals, relative frequencies, or displays. Participants were tested individually.

In each trial, participants were presented with a proportion expressed as a decimal, a relative frequency, or a display of black and white dots (as previously described). This proportion was presented for 3 s. After the proportion was removed, there was a 10-s delay before the response screen was presented. The response screen consisted of 50,000 black dots (as previously described). Participants were asked to manipulate the amount of white dots in the response screen so the proportion of white to total dots equaled the previously presented proportion. Participants manipulated the amount of white dots in each display by pressing number keys on the number pad, as previously described. Participants were not told the exact number of white dots that were added or subtracted by each key press. Instead, the amounts were expressed as "small," "moderate," and "large" amounts. Participants pressed the Enter button to indicate that they believed the proportion of white dots in their display matched the presented proportion.

The sessions consisted of 10 practice trials and 150 experimental trials for each participant. During the experimental trials, participants were given a self-timed rest period after every 32 trials. Each session lasted approximately 50 min. The computer recorded the proportions created by participants.

Results

As in Experiment 1, each participant's data were transformed and plotted on a log-log scale[4] (for an example of a typical

[4] The paired sample *t* test revealed that the regression lines fitted to the transformed data accounted for significantly more variance ($M_{r^2} = 0.63$, $SD_{r^2} = 0.23$) than the regression lines fitted to the raw data ($M_{r^2} = 0.44$, $SD_{r^2} = 0.25$), $t(101) = 10.25$, $p < .001$. In addition to linearizing the data, the log-log transformation permits the data from Experiment 2 to be easily compared with that of Experiment 1.

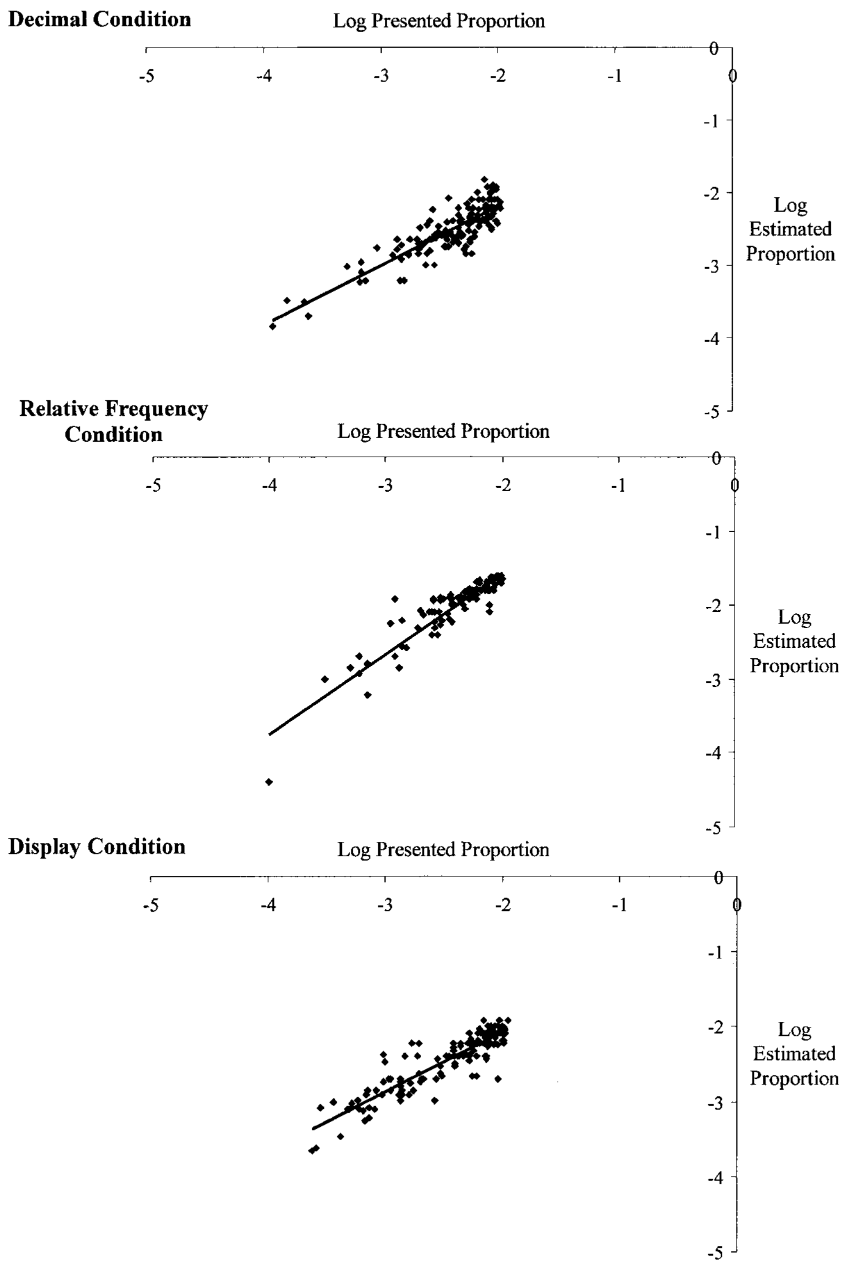

Figure 5. Log-log plots of presented proportion by estimated proportion for a typical participant in the decimal, relative frequency, and display conditions of Experiment 2. Linear functions are fit to the data.

participant in each condition, see Figure 5). A regression line was fit to each participant's data, and the slope, intercept, and R^2 value of each participant's function were used as dependent variables. Ten participants were excluded from the data analysis because their data indicated they were using a strategy that was very different from the majority of the participants. Responses for 4 of these participants were negatively related to the presented proportions. Five of these participants showed no systematic pattern of estimating proportions in relation to the presented proportions and had R^2 values less than .08. Data from 1 participant were excluded because he or she changed estimation strategies during the experiment.[5] Of the remaining 102 participants, 59 were women and 43 were men. To reduce

[5] When viewing the data from this participant, one can see two distinct lines. One line is the result of Trials 1–73, and the second line is the result of Trials 74–150. This finding shows that this participant changed strategies at Trial 74. Because the data could not be characterized well by a linear regression, the data were removed.

Table 2
Means and Standard Deviations of Slopes, Intercepts, and R^2 Values in Experiment 2 by Presentation Type

		Slope		Intercept		R^2	
Presentation type	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Decimal	35	0.796	0.371	−0.375	1.095	0.589	0.282
Relative frequency	34	1.023	0.462	0.270	1.143	0.638	0.212
Display	33	0.783	0.186	−0.481	0.466	0.675	0.193

noise in the data, we did not include the first 35 trials for each participant in the data analysis.[6]

Means and standard deviations for slopes, intercepts, and R^2 values for the presentation types in Experiment 2 are shown in Table 2. A 3 × 2 between-subjects ANOVA for presentation type by gender was performed on each of the three dependent variables (i.e., slopes, intercepts, and R^2 values).

There was a significant main effect of presentation type on slopes, $F(2, 96) = 4.77$, $p = .011$, $MSE = 0.13$ (see Figure 6). A Tukey's post hoc analysis revealed that the slopes of participants who were presented with relative frequencies were higher than the slopes of participants who were presented with decimals ($p < .05$) or displays ($p < .05$). There was no significant difference between the decimal condition and the display condition ($p > .05$). There were no other significant effects on slopes.

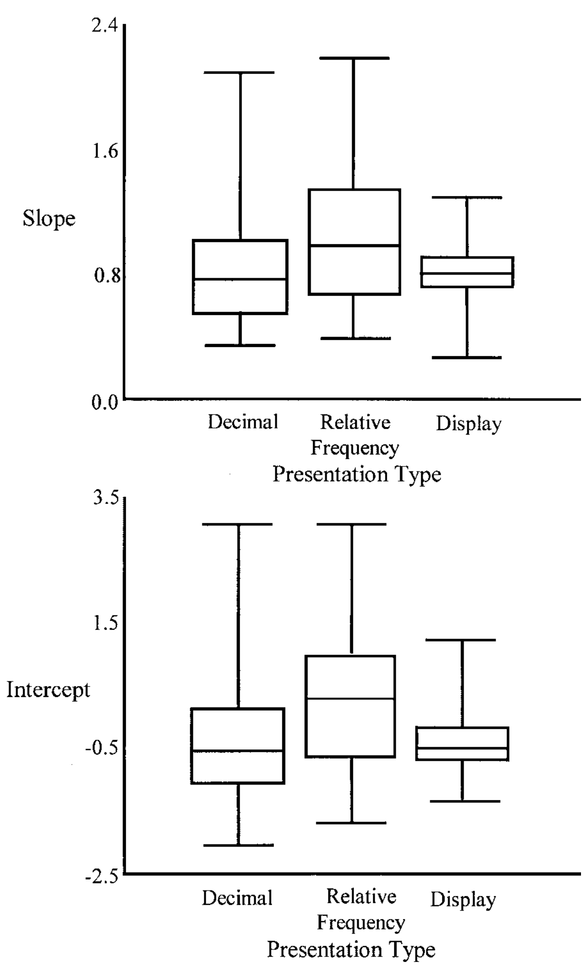

Figure 6. Box plots of the slopes and intercepts in the decimal, relative frequency, and display conditions of Experiment 2.

There was a significant main effect of presentation type on intercepts, $F(2, 96) = 6.08$, $p = .003$, $MSE = 0.93$ (see Figure 6). A Tukey's post hoc analysis showed that the intercepts of participants who were presented with relative frequencies were higher than the intercepts of participants who were presented with decimals ($p < .05$) or displays ($p < .05$). Again, there was no significant difference between participants in the decimal condition and the display condition ($p > .05$). There were no other significant effects on intercepts. There were no significant effects on R^2 values.

To determine if the results of Experiment 2 were the inverse of Experiment 1, we calculated the inverse functions for each participant in Experiment 1. The slopes and intercepts of these inverse functions were compared with the slopes and intercepts of participants in Experiment 2 using an independent *t* test. The means and standard errors for the slopes and intercepts of the decimal and relative frequency conditions in Experiment 2 and the inverse functions of Experiment 1 are presented in Table 3. For the decimal conditions, the slopes of the inverse functions of Experiment 1 were significantly higher than the slopes of Experiment 2, $t(73) = 2.85$, $p = .006$, and the intercepts of the inverse functions of Experiment 1 were significantly lower than the intercepts for Experiment 2, $t(73) = –2.94$, $p = .004$. For the relative frequency conditions, the slopes of the inverse functions of Experiment 1 were significantly higher than the slopes of Experiment 2, $t(72) = 3.55$, $p < .001$, and the intercepts of the inverse functions of Experiment 1 were marginally significantly different from the intercepts for Experiment 2, $t(72) = 1.97$, $p = .052$. Thus, the functions obtained in this experiment were not the inverse of those found in Experiment 1. Although the classical inverse pattern was not shown, there is some evidence that opposing processes are involved in Experiments 1 and 2. Specifically, for both the decimal and relative frequency conditions, the functions from Experiment 2 are curved in opposite directions as the functions from Experi-

[6] By removing the first 35 trials for each participant, 65% of the participants' R^2 values increased. The mean absolute value change for participants whose values increased ($M = .088$, $SD = 0.012$) changed significantly more than for participants whose values decreased ($M = .037$, $SD = 0.004$), $t(101) = 7.60$, $p = .002$. The average R^2 value after the first 35 trials were removed was .62 compared with the previous average R^2 value of .58. When the first 35 trials are included in the analyses, pattern of data does not change.

Table 3
Means and Standard Errors of Slopes and Intercepts for Experiment 2 and the Inverses of Experiment 1 by Condition

Condition	Slope *M*	Slope *SE*	Intercept *M*	Intercept *SE*
Decimal				
Experiment 1 inverse	1.122	0.092	−1.039	0.136
Experiment 2	0.796	0.063	−0.375	0.185
Relative frequency				
Experiment 1 inverse	2.048	0.257	1.443	0.522
Experiment 2	1.023	0.079	0.270	0.196

ment 1. The participants' average functions for each presentation type can be seen in Figure 7.

Discussion

Experiment 2 examined how different presentation types (relative frequency, decimals, and displays) of proportions affected participants' productions of displays representing the proportions. There were several important findings. First, participants estimated proportions differently depending on whether they were presented proportions in relative frequency or decimal format. Second, although all of the participants produced displays that overestimated the presented proportions, participants who were presented with decimals and displays were more accurate than those who were presented with relative frequencies. Third, the relation between the presented proportions and the estimated proportions was again a power function. Fourth, although we had predicted that the results of Experiment 2 would be the inverse of the results of Experiment 1, the proportions in the displays created by participants in Experiment 2 were not the mathematical inverses of the estimations in Experiment 1. Finally, there was no effect of gender.

If participants' estimates were only affected by stimulus and numerical transformation functions, then the results of Experiments 1 and 2 would be inversely related. The fact that the results of Experiments 1 and 2 were not inversely related suggests that participants' estimates are influenced by a response bias, which is revealed by participants' overestimation in both Experiments 1 and 2 (Brooke & MacRae, 1977). This result underscores that participants' biases are revealed only in the combined judgment and production results. Comparisons of the patterns found in Experiments 1 and 2 are discussed in the General Discussion.

The finding that participants' estimates were relatively accurate in the display condition suggests that there is little influence of unidentified effects such as memory. The finding that participants' estimates in the decimal condition were relatively accurate, however, does not suggest that the stimulus and numerical transformation functions for decimals are negligible. Such a conclusion requires the assumption that participants' estimates are only affected by stimulus and numerical transformation functions. Our results refute this assumption. Instead, our results suggest that the response bias offsets the effects of the stimulus and numerical transformation functions in the decimal condition of the production task (see General Discussion).

Experiment 3

In Experiment 3, we assess whether participants' estimates vary as a function of the total number of dots on the screen. If there is an effect of number of dots on $g(\phi)$ within each numerical format condition, the effect can be ascribed to different stimulus transformation functions. Here, we present the participants with a judgment task identical to Experiment 1, with two exceptions: (a) We varied the number of dots on the screen randomly between 500 and 50,000, and (b) we presented only three different proportions (.002, .02, and .2). By assessing these functions for three different proportions, we can determine whether the relation between the

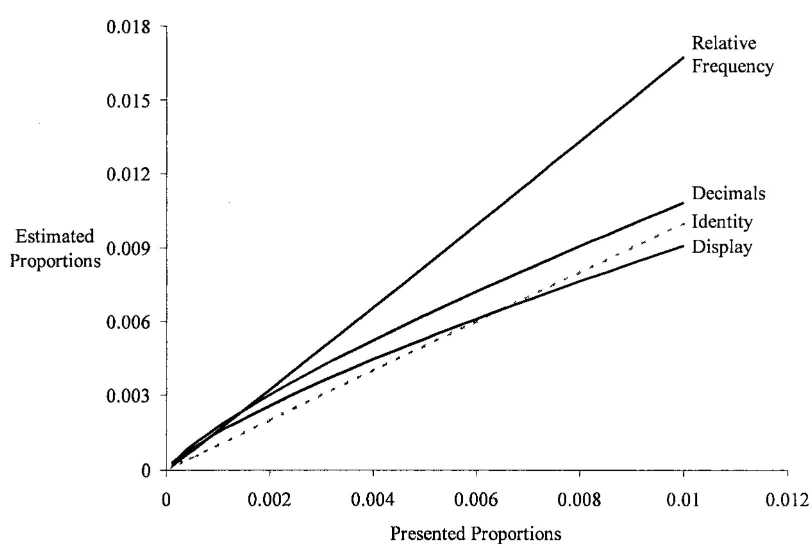

Figure 7. The best fit functions of estimated proportions, or *f*(presented proportions), for participants in the decimal, relative frequency, and display conditions of Experiment 2. These functions were calculated using the average slope and intercept across participants and are plotted on raw scales so the curvature of the functions can be seen. Also included is the identity line (i.e., slope = 1.0).

total number of dots and participants' estimates of low proportions (.02 and .002) is qualitatively different from the relation between the total number of dots and participants' estimates of high proportions (.2). Finally, because there were no effects of gender in Experiments 1 and 2, we dropped gender as a variable.

Method

Participants. Data were collected from 80 introductory psychology students drawn from the same population as in Experiment 1. Volunteers received class credit for participation.

Apparatus. All of the stimuli were presented on a 17-in. (43-cm) VGA color monitor with a 75-Hz refresh rate controlled by a Pentium microcomputer using the DOS operating system. The resolution of the monitor was 1024 × 768.

Stimuli. The displays were identical to Experiment 1 with the following three exceptions. First, the proportion of white to total dots was randomly chosen from one of three values: .002, .02, and .2. Second, the total number of dots on the screen was randomly chosen from a uniform distribution ranging from 500 to 50,000 dots, with the constraint that the total number of dots was evenly divided by the proportion presented. Finally, the display background was gray.

Procedure. Participants were randomly assigned to one of two response type conditions: decimals or relative frequency. The procedure was identical to that of Experiment 1.

Results

Each participant's data were transformed and plotted on a log-log scale. For each participant's data, a regression line assessing the relation between the participant's estimate and the total number of dots on the screen was fit for each presented proportion. The slope, intercept, and R^2 value of each participant's function were used as dependent variables. Importantly, the R^2 values from these functions ($M = .10$, $SD = .126$) were equivalent to functions fit to the untransformed data ($M = .105$, $SD = .12$). The likely reason for this is that the participants' estimates are not well predicted by the total number of dots on the screen. Therefore, we present the analyses from the untransformed data.

Means and standard deviations for slopes, intercepts, and R^2 values for the presentation types in Experiment 3 are shown in Table 4. A 3 × 2 mixed-subjects ANOVA for presented proportion by response type was performed on each of the three dependent variables (i.e., slopes, intercepts, and R^2 values).

Table 4
Means and Standard Deviations of Slopes, Intercepts, and R^2 for Presented Proportion by Response Type for Experiment 3

	Slope		Intercept		R^2	
Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*
Decimal						
.002	0.0000003	0.0000008	.05	.04	.05	.08
.02	0.0000006	0.0000002	.15	.08	.09	.10
.2	0.000005	0.0000006	.41	.14	.20	.15
Relative frequency						
.002	0.000000003	0.000002	.05	.1	.08	.1
.02	−0.0000004	0.0000007	.08	.05	.05	.05
.2	0.000003	0.000003	.30	.17	.12	.13

There was a significant main effect of presentation proportion on slopes, $F(2, 156) = 46.8$, $p < .001$, $MSE = 4.3\text{E-}10$. A Tukey's post hoc analysis revealed that the slopes were higher when participants estimated a proportion of 0.2 than when they estimated a proportion of 0.02 ($p < .05$) or 0.002 ($p < .05$). There was no significant difference between proportions of 0.02 and 0.002 ($p > .05$; see Figure 8). There was a significant main effect of response type on slopes, $F(1, 78) = 6.92$, $p = .01$, $MSE = 6.7\text{E-}11$, such that the slopes of participants who responded using relative frequencies were lower than the slopes of participants who responded using decimals. There was no significant interaction between presented proportion and response type on slopes.

There was a significant main effect of presentation proportion on intercepts, $F(2, 156) = 229.1$, $p < .001$, $MSE = 2.05$. A Tukey's post hoc analysis revealed that the intercepts were higher when participants' estimated a proportion of 0.2 than when they estimated a proportion of 0.02, which in turn were higher than when they estimated a proportion of 0.002. There was a significant main effect of response type on intercepts, $F(1, 78) = 14.24$, $p < .001$, $MSE = 0.023$, such that the intercepts of participants who responded using relative frequencies were lower than the intercepts of participants who responded using decimals. There was a significant interaction between presented proportion and response type on intercepts, $F(2, 156) = 7.26$, $p = .001$, $MSE = .065$. A Tukey's post hoc analysis revealed that, as presented proportion increased, the intercepts of participants in the decimal condition became increasingly larger than the intercepts of participants in the relative frequency conditions.

There was a significant main effect of presentation proportion on R^2, $F(2, 156) = 28.43$, $p < .001$, $MSE = 0.23$. A Tukey's post hoc analysis revealed that the R^2 values were higher when participants were presented with a proportion of 0.2 than when they were presented with a proportion of 0.02 ($p < .05$) or 0.002 ($p < .05$). There was no significant difference between proportions of 0.02 and 0.002 ($p > .05$). There was no significant main effect of response type on R^2. There was significant interaction between presented proportion and response type on R^2, $F(2, 156) = 7.82$, $p < .001$, $MSE = 0.063$. A Tukey's post hoc analysis revealed that the R^2 for the 0.2 presentation proportion in the decimal condition was greater than all other R^2 values.

Discussion

Experiment 3 examined how participants' estimates in a judgment task varied as a function of the total number of dots on the screen. There were two important findings. First, when the presented proportion is .2, the participants' estimates of white to total dots increased with total number of dots for both the decimal and relative frequency conditions. But, when the presented proportion is either .02 or .002, there is little or no effect of the total number of dots on participants' estimates. This finding is reflected in both the slopes and the R^2 values and indicates that participants' estimates are robust over the total number of dots when presented very low proportions (.02 and .002) but not when presented higher proportions (.2). Second, participants treated decimals and relative frequency differently, such that they overestimated when using decimals more than they did when using relative frequency, and this difference increased with larger proportions. This finding is reflected in the intercept data and mimics the major findings of Experiment 1.

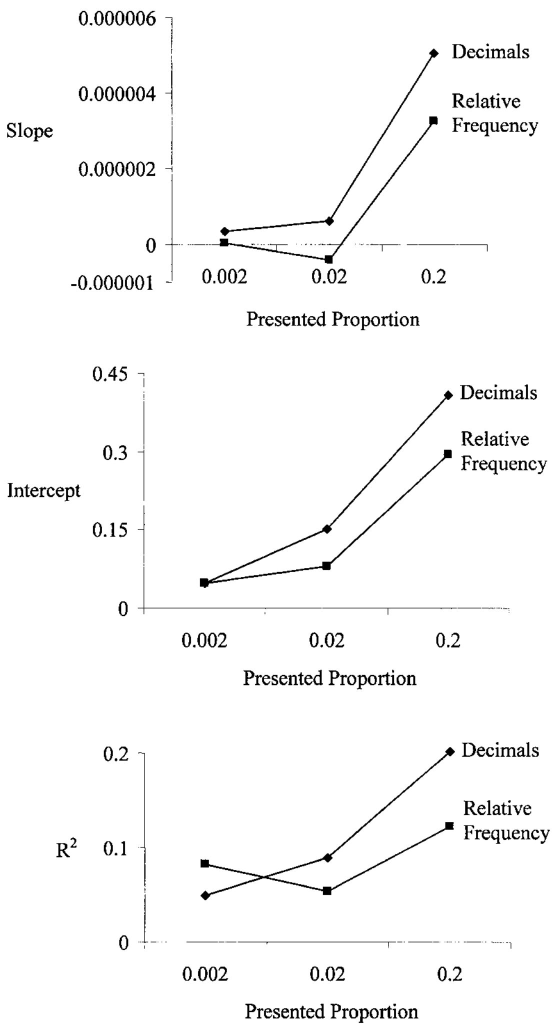

Figure 8. Plots of the slope, intercept, and R^2 values from Experiment 3 for the best fit functions of estimated proportions, or *f*(total number of dots in the display), for each presentation proportion by response type.

Experiment 4

The results of Experiment 3 demonstrated that the biases associated with probability judgment were consistent over total number of dots for low proportions (.002 and .02) but not for relatively high proportions (.2). In Experiment 4, we explored whether probability production is robust over total number of dots. Here, we presented the participants with a production task identical to Experiment 2, with three exceptions: (a) We varied the total number of dots on the response screen randomly between 500 and 50,000, (b) we presented only three different proportions (.002, .02, and .2), and (c) we assessed only the decimal and relative frequency presentation formats (we did not have a display condition).

Method

Participants. Data were collected from 80 introductory psychology students drawn from the same population as in Experiment 1. Volunteers received class credit for participation.

Apparatus. The apparatus were the same as in Experiment 3.

Stimuli. The displays were identical to Experiment 2 with the following three exceptions. First, the proportion of white to total dots was randomly chosen from one of three values: .002, .02, and .2. Second, the total number of dots on the response screen was randomly chosen from a uniform distribution ranging from 500 to 50,000 dots, with the constraint that the total number of dots could be evenly divided by the proportion presented. Finally, the display background was gray.

Procedure. Participants were randomly assigned to one of two presentation type conditions: decimals or relative frequency. The procedure was identical to that of Experiment 2.

Results

Each participant's data were transformed and plotted on a log-log scale. For each participant's data, a regression line assessing the relation between the participant's estimate and the total number of dots on the response screen was fit for each presented proportion. The slope, intercept, and R^2 value of each participant's function were used as dependent variables. The R^2 values from these functions ($M = .01$, $SD = .02$) were equivalent to functions fit to the untransformed data ($M = .01$, $SD = .02$). Again, the likely reason for this is that the participants' estimates are not well predicted by the total number of dots on the response screen. Indeed, the regressions accounted for only 1% of the variance. Thus the intercepts, as opposed to the slopes, will carry most of the information about participants' estimates. Again, we present the analyses from the untransformed data.

Means and standard deviations for intercepts and R^2 values for the presentation types in Experiment 2 are shown in Table 5. A 3×2 mixed-subjects ANOVA for presented proportion (.2, .02, or .002) by presentation type (decimal vs. relative frequency) was performed on each of the three dependent variables (i.e., slopes, intercepts, and R^2 values).

There were no significant main effects or interaction on slopes. Furthermore, the slopes were not significantly different from zero, $t(79) = 0.4$, *ns*.

There was a significant main effect of presentation proportion on intercepts, $F(2, 156) = 41.01$, $p < .001$, $MSE = 0.085$. A Tukey's post hoc analysis revealed that the intercepts were higher when participants were presented with a proportion of .2 than when they were presented with a proportion of .02 ($p < .05$) or .002 ($p < .05$). There was no significant difference between proportions of .02 and .002 ($p > .05$). There was no significant main effect of presentation type on intercepts. There was a signif-

Table 5
Means and Standard Deviations of Intercepts and R^2 for Presented Proportion by Presentation Type for Experiment 4

Condition	Intercept *M*	Intercept *SD*	R^2 *M*	R^2 *SD*
Decimal				
.002	.05	.08	.01	.01
.02	.05	.07	.01	.01
.2	.08	.09	.02	.02
Relative frequency				
.002	.06	.15	.01	.01
.02	.08	.15	.01	.01
.2	.15	.13	.01	.01

icant interaction between presented proportion and response type on intercepts, $F(2, 156) = 4.33$, $p = .01$, $MSE = 0.009$. A Tukey's post hoc analysis revealed that, as presented proportion increased, the intercepts of participants in the relative frequency condition became increasingly larger than the intercepts of participants in the decimal conditions.

There was a significant main effect of presentation proportion on R^2, $F(2, 156) = 3.5$, $p = .03$, $MSE = 0.0009$. A Tukey's post hoc analysis revealed that the R^2 values were higher when participants were presented with a proportion of 0.2 than when they were presented with a proportion of 0.02 ($p < .05$) or 0.002 ($p < .05$). There was no significant difference between proportions of 0.02 and 0.002 ($p > .05$). There was no significant main effect of presentation type on R^2. There was a significant interaction between presented proportion and presentation type on R^2, $F(2, 156) = 3.53$, $p = .03$, $MSE = 0.0009$. A Tukey's post hoc analysis revealed that the R^2 of the 0.2 presentation proportion in the decimal condition was greater than all other R^2 values.

Discussion

Experiment 4 examined how participants' estimates in a production task varied as a function of the total number of dots on the response screen. There were two important findings. First, there was no effect of total number of dots on participants' estimates. This finding is reflected in both the slopes and the R^2 values and indicates that participants' estimates in a production task are robust over the total number of dots on the response screen. Second, participants treated decimals and relative frequency differently, such that they overestimated when making displays from relative frequencies more than they did when making displays from decimals, and this difference increased with larger proportions. This finding is reflected in the intercept data, and it mimics the major findings of Experiment 2.

Experiment 5

Experiment 5 is an attempt to determine whether relative frequencies and decimals share a single quantity representation that mediates all judgment and production processes (e.g., McCloskey, 1992). If relative frequencies and decimals share a single quantity representation that mediates judgment and production processes, then the bias associated with the conversion between the two numerical formats should be predicted by Equation 8. If relative frequencies and decimals do not share a single quantity representation that mediates judgment and production processes, then the bias associated with the conversion between the two numerical formats should be unpredictable. To test these predictions, in Experiment 5 we asked participants to convert relative frequencies to decimals and vice versa.

Method

Participants. Data were collected from 112 participants drawn from the same population as in Experiment 1. Eighty-two participants were women and 30 participants were men.

Procedure. Participants were randomly assigned to one of two conditions: relative frequency or decimals. Participants were tested individually. The same rooms and computers used in the previous experiments were used here.

In one condition, participants were presented a proportion expressed in decimal format and were asked to reexpress that proportion in relative frequency format. In the other condition, participants were presented a proportion expressed in relative frequency format and were asked to reexpress that proportion in decimal format. The presented proportions were identical to those of the decimal and relative frequency conditions in Experiment 2. The response screen immediately followed the presented proportion. The response screens were identical to those used in Experiment 1. Each session consisted of 150 experimental trials (there were no practice trials). Participants were given a rest period after every 42 trials. Each session lasted approximately 20 min. The computer recorded participants' responses.

Results

Each participant's data were transformed and plotted on a log-log scale. Because many participants' data produced multiple parallel lines (i.e., lines of equal slope but different intercepts), only the participants' slopes were used as the dependent variable. The intercepts were dropped as a dependent variable because they varied within sessions. R^2 was dropped as a dependent variable because the participants' data were extraordinarily linear. A typical participant's data are presented in Figure 9.

Data from 27 participants were excluded from the data analysis. Two participants did not complete the experiment. Eighteen participants used six different responses for all of the trials (suggesting that they did not follow instructions). Seven participants showed no patterns in their responses in relation to the presented proportions. Of the remaining 85 participants, 40 were in the decimal condition and 45 were in the relative frequency condition.

There were no significant differences between the slopes of participants who were presented with decimals ($M = -0.72$, $SD = 1.66$) and the slopes of participants who were presented with relative frequencies ($M = -0.58$, $SD = 0.95$), $t(84) = -0.59$, $p = .55$. Examination of the participants' data revealed that participants used two different strategies to produce their responses. One

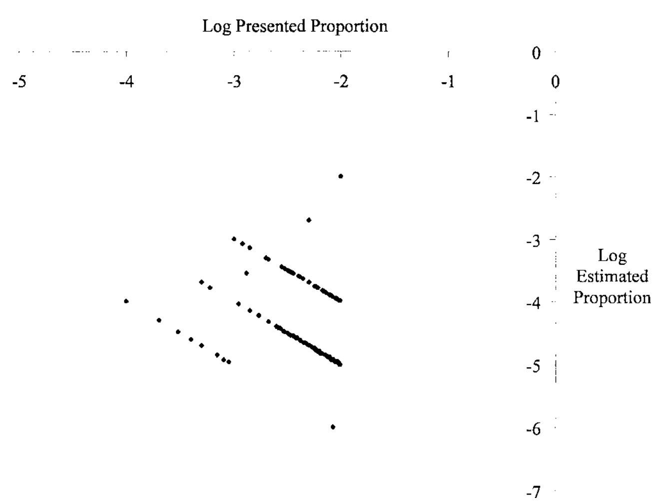

Figure 9. Log-log plot of a typical participant in Experiment 5 who incorrectly converted relative frequencies to decimals by using the numbers in the presented proportion when making his or her response (see text). The multiple parallel lines are a result of the participant shifting the decimal point in his or her responses.

strategy, used by 15 out of the 85 participants who showed a pattern of responding, was to attempt to mathematically convert decimals into relative frequencies or vice versa. Of these 15 participants, 7 were presented with decimals and 8 were presented with relative frequencies. Participants using this strategy estimated proportions that were either fairly accurate or differed from the presented proportion by one or two orders of magnitudes. The mean slope of the lines produced by using this strategy was 1.09 ($SD = 0.57$).

In contrast, 70 out of the 85 participants who showed a pattern of responding produced lines with *negative* slopes. Of these 70 participants, 33 were presented with decimals and 37 were presented with relative frequencies. The mean slope of the lines produced by using this strategy was –1.02 ($SD = 0.83$). Participants showing this pattern of responding did not retain the ordinal structure of the numerical scales.

Discussion

The purpose of Experiment 5 was to assess the biases associated with converting relative frequencies to decimals and vice versa. The results showed that (a) about 24% of participants did not complete the experiment properly or showed no relation between relative frequencies and decimals, (b) about 13% of participants accurately converted between the two numerical formats, and (c) about 63% of participants used a strategy that produced a negative relation between decimals and relative frequencies. These biases are not predictable from the biases present in Experiments 1 and 2. The data from Experiments 1 and 2 show that participants retained the correct ordinal relation between both numerical formats and the visual displays, but the accuracy and reliability of participants' estimates were imperfect. The data from Experiment 5 suggest that the vast majority of participants could not retain the correct ordinal relation when converting between relative frequencies and decimals.

When we compared the presented proportions with the estimated proportions, we found that the participants whose data showed a negative relation used the numbers in the presented proportions in their responses. For example, if participants were presented with ".0074," they may have incorrectly responded that that number is equal to "1 in 7,400." Likewise if participants were presented with "1 in 200," they may have incorrectly responded that that number is equal to ".02." This strategy is likely an overgeneralization of a few instances where this works. For example, "1 in 100" is represented by .01, and "1 in 1,000" is represented by .001. If students incorrectly generalized this pattern to the rest of the data, then their slopes would equal –1.0. This is consistent with the observed data. Often, participants' graphs showed several lines of the same slope but different intercept (see Figure 9). These multiple lines were the result of changing the magnitude of their responses. For example, a participant may have converted .075 to "1 in 750" in one trial and converted .085 to "1 in 8,500" in another trial. These data, together with mathematically correct conversion computed by a minority of participants, strongly suggest that participants used rule-based strategies to transform relative frequencies to decimals (and vice versa) rather than accessing a common abstract representation.

General Discussion

Summary of Major Results

The results of the present experiments provide several insights about how people estimate very low proportions. In Experiment 1, participants were shown proportions expressed as displays of dots and were asked to estimate the proportions as relative frequencies or decimals. The results of Experiment 1 demonstrated that (a) the relation between estimated and presented proportions was a power function, (b) participants overestimated very low proportions, and (c) participants estimated proportions differently depending on whether they labeled the proportions as relative frequencies or decimals. In Experiment 2, participants were shown proportions expressed as relative frequencies, decimals, or displays and were asked to create a display of dots that matched the presented proportions. The results of Experiment 2 demonstrated that (a) the relation between estimated and presented proportions was a power function, (b) participants overestimated very low proportions, (c) participants estimated proportions differently depending on whether they were presented proportions in relative frequency or decimal format, and (d) the estimation patterns found in Experiment 2 were not inversely related to the patterns in Experiment 1. In both Experiments 1 and 2, there was no effect of gender. In Experiments 3 and 4, we varied the total number of dots in the judgment and production tasks. The results of Experiments 3 and 4 demonstrated that there is little or no effect of the total number of dots either in the presentation screen or the response screen on the participants' estimates. The one exception was when participants were asked to estimate proportions of .2 in the judgment task. In that instance, there was an effect of the total number of dots. Finally, in Experiment 5, participants were asked to convert relative frequencies to decimals and vice versa. The results of Experiment 5 demonstrated that (a) the relation between estimated and presented proportions was a power function and (b) the majority of participants produced patterns showing a negative relation between relative frequencies and decimals.

Overall, these experiments demonstrated that participants were inaccurate when estimating very low proportions, and participants' estimates were generally robust over the total number of dots. Below we discuss the various biases associated with our participants' responses.

Biases in Labeling Perceived Probability

Our data allow us to detect three types of biases associated with our participants' responses: a *response bias*, a *perceptual bias*, and a *numerical transformation bias*. To assess these biases, we compared our participants' responses across conditions and experiments. The presence of each bias is indicated by a particular pattern of response. Unfortunately, although our data allow us to identify the presence and general structure of these biases, they do not allow us to precisely quantify each bias. This limitation exists because several biases are present in the data, and each bias may take the form of a power function. Therefore, there are too many free parameters to mathematically identify each bias.

Response bias. A response bias occurs when participants' estimates are influenced by a partiality toward certain response options. A response bias is revealed if participants' estimates in the judgment and production tasks are not inversely related (see Brooke & MacRae, 1977). Our data exhibited this pattern. The response bias revealed in the present experiments takes the form of a consistent overestimation in both judgment and production tasks (Experiments 1 and 2), indicating a bias against very low proportions (under .01).

There are several mathematical forms that a response bias may take. One relatively simple form is that the output of $g(\phi)$ and $g^{-1}(\#)$ are both transformed in the same way,

$$R_{\#} = f_R(g(\phi)) \quad \text{and} \quad R_{\phi} = f_R(g^{-1}(\#)), \tag{10}$$

where f_R is a response function and $R_{\#}$ and R_{ϕ} are the numerical and stimulus responses, respectively. In this form, f_R is mathematically identified:

$$f_R^{-1}(R_{\#}) = f_R(R_{\phi}). \tag{11}$$

Furthermore, this model makes the prediction that when the inverse response transformation, f_R^{-1}, is applied to the judgment and production responses, the resulting functions should be the inverse of one and other,

$$f_R^{-1}(R_{\#}) = g(\phi) \quad \text{and} \quad f_R^{-1}(R_{\phi}) = g^{-1}(\#). \tag{12}$$

We calculated f_R using Equation 11, assuming f_R takes the form of a power function. The resulting response functions for the decimal, $\log(y) = -0.5273 + 0.7812 * \log(x)$, and relative frequency, $\log(y) = 0.5744 + 1.21 * \log(x)$, conditions failed to satisfy the prediction of Equation 12. Therefore, this is an unlikely model for the response bias.

A more likely model of response bias assumes that participants overestimate because they have difficulty anchoring their responses. One such model accounts for overestimation by adding a constant (a) to the input of the numerical transformation function. Therefore in the judgment task,

$$R_{\#} = f_N(f_S(\phi) + a), \tag{13}$$

and in the production task,

$$R_{\phi} = f_S^{-1}(f_N(\# + a)). \tag{14}$$

Unfortunately, because Equations 13 and 14 are not identified, the model cannot be tested.[7]

Regardless of the form that the response function takes, the present experiments are the first psychophysical experiments to demonstrate a response bias against very low proportions. Although overestimation of proportions under 0.5 in the judgment task is a very robust finding (e.g., Brooke & MacRae, 1977; Hollands & Dyre, 2000; Spence, 1990; Varey et al., 1990), only participants in Brooke and MacRae's experiments completed both the judgment and production tasks. Brooke and MacRae found no evidence for a response bias.[8] The difference between the findings of the present experiments and the findings of Brooke and MacRae's experiments may be due to the different ranges of proportions used. Recall that the lowest proportion presented in Brooke and MacRae's experiment was .10, so their participants never had an opportunity to exhibit a response bias against proportions under .01.

Perceptual bias. A perceptual bias refers to a propensity to estimate a proportion in a certain direction because of a misperception of the stimulus (see Hollands & Dyre, 2000, for a discussion of perceptual bias in proportion estimation). Brooke and MacRae (1977) stated that if participants have a perceptual bias when estimating proportions, inverse tasks (judgment and production) should produce inverse patterns of estimations. This is only true, however, when (a) $f_N = 1$, and (b) there is no evidence of a response bias.

To assess the presence of a perceptual bias without arbitrarily fixing $f_N = 1$ or in the presence of a response bias, one must hold the numerical transformation function constant and vary the physical stimulus. If $g(\phi)$ and $g^{-1}(\#)$ vary with the physical stimulus, then the variation must be the result of different stimulus transformation functions, and therefore different perceptual biases. In Experiments 3 and 4, we presented participants with displays that varied in their total numbers of dots. By comparing participants' responses within each numerical format condition, we fix the numerical transformation function. We can therefore ascribe any difference between the patterns of responses across displays to different stimulus transformation functions.

The results of Experiments 3 and 4 indicate that, in most conditions, there was little effect of the total number of dots on participants' responses. The one exception to this finding was the .2 display in the judgment task for both the relative frequency and decimal conditions. In these conditions, $g(\phi)$ varied with the total number of dots. Thus, for the relatively large proportion of .2, there is evidence for a distinct perceptual bias.

The presence of a perceptual bias of the .2 display is not unexpected given that perceptions of virtually every physical stimulus from light to electric shock are biased (e.g., Marks, 1974; Stevens, 1986). It is noteworthy, however, that no direct evidence was found for perceptual biases associated with displays representing proportions below 2%. Because perceptual biases are detected only in relation to one and other, the lack of evidence for a perceptual bias suggests a constancy of bias across displays rather than a lack of bias. Thus, our participants' perceptions of proportions below 2% were relatively stable, whereas those over 2% were relatively volatile. The source of this volatility should be explored.

Numerical transformation bias. A numerical transformation bias occurs when one imperfectly converts a stimulus into a numerical label and vice versa. For example, when participants are asked to describe a perception in terms of a number, they must transform their perception into a number. If this transformation is different for each numerical format (e.g., relative frequency, decimals, and magnitude estimates), participants have a numerical transformation bias.

To assess the presence of a numerical transformation bias, one must hold the stimulus transformation function constant and vary the numerical format. If $g(\phi)$ and $g^{-1}(\#)$ vary with the numerical format, then the variation must be the result of different numerical transformation functions, and therefore different numerical transformation biases. In the present experiments, participants in both

[7] Although we have only identified two forms that the response function may take, there are innumerable other possible forms. Most, if not all, of these forms are not identified and therefore cannot be tested with the present data.

[8] The perceptual bias found in Brooke and MacRae's (1977) study should be accepted only with caution. The participants in Brooke and MacRae's study always completed the judgment phase before the production phase. Therefore, in Brooke and MacRae's study, the results of the production phase were confounded by the participants' experiences in the judgment phase. Because we used different participants in Experiments 1 and 2, this was not an issue in our experiments.

the decimal and relative frequency conditions were presented with the same stimuli. Therefore, the stimulus transformation function was fixed, and any difference between the patterns of responses for the two numerical format conditions can be ascribed to different numerical transformation functions.

To assess whether the numerical transformation for relative frequency is identical to that for decimals, we examined the slopes of each condition. Because the power law can describe our data, the slope of the log-log transformed data carries the information about how the participants' reported magnitude varies with the magnitude of the presented stimulus (Stevens, 1986). Therefore, if the numerical transformation function for relative frequency is identical to that for decimals, one would expect to find similar slopes in our participants' data. If the slopes are different, participants have a numerical transformation bias.

In both the judgment task (Experiment 1) and production task (Experiment 2), participants' slopes in the decimal and relative frequency conditions were significantly different. In fact, the data from the decimal condition curved in the opposite direction as that from the relative frequency condition (as indicated by slopes above and below 1). This demonstrates the existence of a numerical transformation bias.

The discovery of a numerical transformation bias augments our understanding of biases in proportion estimation. Specifically, our data indicate that the pattern of the estimation error is in part dependent on the format of the numerical labels used when collecting estimates. Although this finding has little impact on the validity of models of estimation bias, such as Hollands and Dyre's (2000) cyclical power model, it does suggest that the critical exponent estimated using those models would depend on the format of the numerical label requested of the participant. Further implications of the numerical transformation bias are discussed below (see the *Implications* section below).

Robustness of biases. Experiments 1 and 2 show no gender-specific response biases (i.e., our findings were robust over gender). This finding is important, because it helps validate a long-held but rarely tested assumption characteristic of any experiment that uses probability estimates to assess cognitive states such as beliefs and attitudes (e.g., when participants are asked to estimate their probability of contracting HIV; D. J. Cohen & Bruce, 1997). Researchers assume that gender effects in such experiments do not simply reflect differences in how men and women use and interpret numbers (i.e., not an effect of the response class). This assumption is not risk free given that gender-specific effects are sometimes found in spatial and mathematical reasoning tasks (for a review, see Hyde & McKinley, 1997). Because the results of Experiments 1 and 2 show no difference in how male and female participants interpret and use very low probabilities, they support the assumption that gender effects in experiments that use probability estimates to assess cognitive states are not simply an effect of the response class.

Experiments 3 and 4 show that participants' estimates are robust over the total number of dots for proportions of .002 and .02 in both the judgment and production tasks. For the proportion of .2, participants' estimates were robust over the total number of dots only in the production task. In the .2 proportion condition of the judgment task, participants' estimates increased as the total number of dots increase. These findings suggest that biases found in Experiments 1 and 2 (which measured estimates of proportions less than .01) would be present regardless of the number of dots in the visual display. In contrast, any biases present for proportions greater than .02 may depend on the number of dots in the visual display. This provides some evidence that participants treat relatively large proportions differently than they treat very low proportions (see the next section, *On Very Low Proportions and Reference Points*). This may contribute to why the biases found in the present series of experiments are different from those found when participants estimate proportions greater than .01 (e.g., Hollands & Dyre, 2000; Spence, 1990).

On Very Low Proportions and Reference Points

The results of Experiments 1 and 2 suggest that the biases associated with estimating very low proportions are different from those associated with estimating proportions between 1% and 99%. Previous research has demonstrated that participants produce an inverse ogival error pattern when estimating proportions between 1% and 99% (e.g., Hollands & Dyre, 2000). Furthermore, Hollands and Dyre have shown that much of that data can be well described by their cyclical power model. In contrast, (a) our data are best described by Stevens's Power Law (Stevens, 1986), and (b) our participants exhibited a response bias against very low proportions.

The fundamental difference between our data and that of Spence (1990) and Hollands and Dyre (2000) is that our participants' failed to adopt an upper reference point. If participants do not adopt an upper reference point, both the cyclical power model and the power model will fail. However, because Stevens's Power Law does not incorporate an upper reference point, it fits our data well.

Our participants' failure to adopt an upper reference point is likely a function of the quantities presented in our experiments. Spence (1990) proposed that participants naturally adopt 1.0 as an upper reference point. The natural reference point of 1.0, however, may have been too distant from the quantities we presented to be used effectively by our participants. Recall that the largest quantity presented in our experiments was 0.01. Furthermore, although Hollands and Dyre (2000) demonstrated that participants will adopt other reference points, these reference points must be salient (e.g., visible marks on a glass). In our experiments, we provide no salient reference points. Thus, our participants may not have adopted an upper reference point because no natural or salient reference point was accessible.

Reference points are critical to the accurate estimation of proportions because reference points act to correct for over- and underestimation by "pulling" participants' estimates toward accuracy. In the absence of a salient or natural reference point, participants' estimates will continuously deviate from accuracy. Consequently, our participants' failure to adopt an upper reference point resulted in functions that cannot be generalized beyond the proportions presented. For example, the function predicting male participants' responses in the decimal judgment task is log(estimated proportion) $=$ 0.917 $+$ 1.231 $*$ log(presented proportion). Although this equation accurately describes male participants' estimates of very low proportions, it provides absurd predictions for larger proportions (e.g., this function predicts that men will estimate "0.5" as "3.519").

Finally, it is worth noting that familiarity with quantities may influence the adoption of reference points. That is, the more familiar one is with a quantity, the more likely that quantity can act as a reference point. The magnitude of our participants' errors suggests that our participants were not even familiar with the quantities near the presented proportions. This finding is not unexpected because very low proportions are infrequently experienced. Therefore, the quantity representations associated with very low proportions may be both inaccurate and noisy. Indeed, the limited research assessing people's understanding of very low proportions suggests that people have extreme difficulty interpreting these values (e.g., Kaplan, Hammel, & Schimmel, 1985; Rothman & Kiviniemi, 1999).

Our hypothesis that participants failed to adopt an upper reference point because no reference point was accessible leads to the prediction that one can increase the accuracy of participants' estimates by making an upper reference point accessible. One can make an upper reference point accessible by (a) providing a salient reference point similar to those used by Hollands and Dyre (2000), (b) expanding the quantities assessed to include or be relatively close to a natural reference point, or (c) increasing participants' familiarity with at least one quantity in the range assessed. It should be noted, however, that very low proportions are rarely communicated with salient reference points (Environmental Protection Agency, 1991). Our results suggest that this lack of salient reference points may explain the difficulty associated with effectively communicating low-risk events.

In sum, estimates of very low proportions appear to be different from estimates of proportions between 1% and 99%. This difference may be a function of participants' inability to adopt an upper reference point. Further research is needed to determine whether this finding is a result of the novelty of the proportions presented or whether it is specific to very low proportions.

Numerical Representation

The results of Experiments 1, 2, and 5 provide the information necessary to assess the validity of the McCloskey model and the Gonzalez and Kolers model. McCloskey (1992) proposed that all comprehension, production, and computational processes are mediated by a single quantity representation. In contrast, Gonzalez and Kolers (1982) proposed that each numerical format is associated with a unique quantity representation. The research that addresses this question has centered on assessing the similarity of number comparison processes across numerical formats (for reviews, see Dehaene, 1992; McCloskey, 1992) rather than assessing participants' intuitive senses of amount associated with each numerical format. In the present experiments, we assessed how participants' intuitive senses of amount associated with each numerical format relate to how participants convert between numerical formats. The data from Experiments 1 and 2 show that participants retained the correct ordinal relation between both numerical formats and the visual displays, whereas the data from Experiment 5 reveal that participants could not retain the correct ordinal relation when converting between the two numerical formats.

The McCloskey model of numerical processing predicts that the bias associated with converting Format A to Format B would equal the combined biases associated with converting Format A to a quantity representation and converting that quantity representation to Format B (Equation 8). The data from Experiment 5 refute this prediction. Instead, the data suggest that the majority of participants used an imperfect rule-based strategy to convert between the two formats. Furthermore, these participants were apparently unaware that their rule-based strategy reversed the relation between the two formats. Thus our data strongly suggest that a single quantity representation did not mediate the conversion between the two formats.

Gonzalez and Kolers's (1982) model, in contrast, anticipates our results. Gonzalez and Kolers hypothesized that there exist multiple, incompatible quantity representations and the incompatibility may "make it impossible to translate one symbol into another" (p. 318). The difficulty translating between numerical formats results from participants' inability to rely on a common quantity representation or even compare quantity representations. Consequently, participants must use a nonanalog process to convert between numerical formats. Furthermore, because no quantity representation is involved, participants will be unaware of inconsistencies between the output of the nonanalog strategy and the quantity representation associated with each numerical format. Our data mirror these predictions and therefore support the hypothesis that relative frequencies and decimals are represented in the human brain by different, incompatible quantity representations.

The source of the difference between our data and the data supporting the McCloskey model may be the quantities used. Most research investigating quantity representations has assessed verbal versus Arabic symbols of integers (for reviews, see Ashcroft, 1992; Dehaene, 1992, 1997; McCloskey, 1992). Because verbal and Arabic symbols of integers are experienced daily, the high familiarity of these symbols may facilitate the use of a shared quantity representation. In contrast, the present experiment assessed two forms of Arabic notation of very low proportions. Because both the symbols of very low proportions and the actual proportions are experienced much less frequently, the quantity representations associated with each numerical format may not have integrated into a single quantity representation. Thus, it may be that the relation between familiarity and quantity representations is critical and requires further exploration.

Implications

In the present article, we show that people have perceptual, response, and numerical transformation biases when estimating very low proportions using relative frequency and decimal formats. The perceptual bias revealed in the present experiments is a function of the participants' perception of the visual displays we used to represent very low proportions. Although this perceptual bias generalized over a large range of the number of dots in the display, the degree to which it generalizes to other nonverbal representations of amount still needs to be explored.[9] However, because the response and numerical transformation biases revealed in the present experiments are theoretically independent of the nonverbal stimulus used to represent amount (i.e., the visual displays), one expects them to generalize to all nonverbal stimuli used to represent amount. These biases have implications to all areas of psychological research that use numerical labels as stimuli or responses.

Of the biases discussed, the numerical transformation bias may have the most serious practical implications. First, because partic-

[9] However, consistent error patterns have been demonstrated across stimuli and perceptual dimension for proportions between .01 and .99 (for a discussion, see Hollands & Dyre, 2000).

ipants use relative frequencies differently than they use decimals, the magnitude of responses from experiments that use one numerical format cannot be directly compared with those that use a different numerical format. However, because response types were ordinally related to the proportions in the visual displays of Experiments 1 and 2, researchers may be able to compare the ordinal relations between conditions. Second, because participants had a consistent bias for each numerical format, researchers can compare the magnitudes of responses between experiments that use the same numerical format. Third, because the participants' psychophysical functions for each numerical format were consistent, these functions may be used to aid researchers in understanding the meaning of participants' estimates of very low probabilities. For example, when participants are asked to describe the risks associated with a very low probability event, any large deviations from these functions may indicate the presence of an effect above and beyond the biases described here.

The numerical transformation bias also has several serious implications beyond those related to relative frequencies and decimals. N. Schwarz (1999) reviewed the psychological literature and provided substantial evidence that response format affects the magnitude of participants' responses in many instances. N. Schwarz hypothesized that many response effects are a result of participants using the response alternatives to clarify the question asked. That is, different response alternatives suggest different interpretations of the same question. Although this is likely the case in many instances, we hypothesize that this is not the case in the present experiments because our task was relatively unambiguous. Instead, we hypothesize that the numerical transformation bias found in the present experiments arises because people associate numerical labels in decimal format with different quantity representations than they do numerical labels in relative frequency format. If this is the case, researchers must be aware of at least two different sources of response effects: (a) the effects due to the information contained in the response format (as suggested by N. Schwarz) and (b) the participants' understanding and use of the numbers themselves (as our data suggest).

One area of psychology in which participants' understanding and use of numbers is critical is that of psychophysics. Perceptual psychologists have used magnitude estimation procedures to determine the relation between physical stimuli and the psychological perception of those stimuli (for reviews, see Gescheider, 1988; Marks, 1974; Marks & Algom, 1998; Stevens, 1956, 1986). The data from magnitude estimation procedures have reliably exhibited a power function. The characteristic exponent, which describes the curvature of the raw data, has been used as the measure of the relation between physical stimuli and the psychological perception of those stimuli. Experiments 1 and 2 demonstrate that estimates using relative frequency and decimal formats also exhibit a power function, but the characteristic exponent was a function of the participant's response format. Specifically, participants' responses in the relative frequency condition consistently exhibited an opposite curvature than participants' responses in the decimal condition. Because magnitude estimates are another form of numerical labeling, there is no a priori reason to believe that magnitude estimates are immune to this numerical transformation bias. If magnitude estimates are susceptible to numerical transformation bias, then interpretation of the characteristic exponent becomes difficult.

Finally, it is worth repeating that the effect of participants' understanding and use of numbers on the data they produce, and therefore the conclusions researchers draw, may be substantial. This point is illustrated in the work of Gigerenzer and his colleagues (e.g., Gigerenzer & Hoffrage, 1995; Seldmeier & Gigerenzer, 2001). Recall that Gigerenzer and his colleagues demonstrated that participants engage in accurate Bayesian reasoning when amount information is expressed as relative frequencies but fail at Bayesian reasoning when amount information is expressed in decimal format. It is of great interest to determine how numerical format interacts with people's other mathematical and reasoning abilities.

Conclusion

In this article, we presented a theoretical and experimental framework for assessing the biases associated with how people interpret numbers, and we showed how this framework could reveal details concerning how quantity information is represented in the human brain. We reported five experiments that assessed how people interpret and represent relative frequencies and decimals that denote very low proportions (i.e., proportions below 1%). Our participants demonstrated perceptual, response, and numerical transformation biases when making their estimates. Furthermore, our data suggest that relative frequencies and decimals are associated with different quantity representations. Our data, therefore, cast doubt on the assumption of numerical equivalence. Without the assumption of numerical equivalence, it becomes essential to determine how numerical format interacts with people's mathematical and reasoning abilities.

References

Ashcroft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition, 44*, 75–106.

Ashcroft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General, 130*, 224–237.

Begg, I. (1974). Estimation of word frequency in continuous and discrete tasks. *Journal of Experimental Psychology, 102*, 1046–1052.

Brooke, J. B., & MacRae, A. W. (1977). Error pattern in the judgment and production of numerical proportions. *Perception & Psychophysics, 21*, 336–340.

Buckley, P. B., & Gillman, C. B. (1974). Comparison of digits and dot patterns. *Journal of Experimental Psychology, 103*, 1131–1136.

Campbell, J. I. D. (1994). Architectures for numerical cognition. *Cognition, 53*, 1–44.

Campbell, J. I. D, & Xue, Q. (2001). Cognitive arithmetic across cultures. *Journal of Experimental Psychology: General, 130*, 299–315.

Cohen, D. J., & Bruce, K. E. (1997). Sex and mortality: Real risk and perceived vulnerability. *Journal of Sex Research, 34*, 1–13.

Cohen, L., Dahaene, S., Chochon, F., Lehericy, S., & Naccache, L. (2000). Language calculation within the parietal lobe: A combined cognitive, anatomical, and fMRI study. *Neuropsychologia, 38*, 1426–1440.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition, 44*, 1–42.

Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York: Oxford University Press.

Dehaene, S., & Akhavein, R. (1995). Attention, automaticity, and levels of representation in number processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 314–326.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science, 284*, 971–974.

Dehaene, S., Tzourio, N., Frak, V., Raynaud, L., Cohen, L., Mehler, J., & Mazoyer, B. (1996). Cerebral activations during number multiplication and comparison: A PET study. *Neuropsychologia, 34*, 1097–1106.

Environmental Protection Agency. (1991, October). *Environmental risk: Your guide to analyzing and reducing risk* (U.S. EPA Region 5, Publication Number 905/9-91/017). Cincinnati, OH: National Service Center for Environmental Publications.

Gescheider, G. A. (1988). Psychophysical scaling. *Annual Review of Psychology, 39*, 169–200.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102*, 584–704.

Gladis, M. M., Michela, J. L., Walter, H. J., & Vaughan, R. D. (1992). High school students' perceptions of AIDS risk: Realistic appraisal or motivated denial? *Health Psychology, 11,* 307–316.

Gonzalez, E. G., & Kolers, P. A. (1982). Mental manipulation of arithmetic symbols. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 308–319.

Hansen, W. B., Hahn, G. L., & Wolkenstein, B. H. (1990). Perceived personal immunity: Beliefs about susceptibility to AIDS. *Journal of Sex Research, 27,* 622–628.

Hollands, J. G., & Dyre, B. P. (2000). Bias in proportion judgments: The cyclical power model. *Psychological Review, 107*, 500–524.

Hyde, J. S., & McKinley, N. M. (1997). Gender differences in cognition: Results from meta-analyses. In P. J. Caplan, M. Crawford, J. S. Hyde, & J. T. E. Richardson (Eds.), *Gender differences in human cognition* (pp. 30–51). New York: Oxford University Press.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*, 430–454.

Kaplan, R. M., Hammel, B., & Schimmel, L. E. (1985). Patient information processing and the decision to accept treatment. *Journal of Social Behavior and Personality, 1,* 113–120.

Marks, L. E. (1974). *Sensory processes*. New York: Academic Press.

Marks, L. E., & Algom, D. (1998). Psychological scaling. In M. H. Birnbaum (Ed.), *Measurement, judgement, and decision making* (pp. 81–178). New York: Academic Press.

McCloskey, M. (1992). Cognitive mechanisms in numerical processing: Evidence from acquired dyscalculia. *Cognition, 44*, 107–157.

McCloskey, M., Caramazza, A., & Basili, A. G. (1985). Cognitive mechanisms in number processing and calculation: Evidence from dyscalculia. *Brain and Cognition, 4*, 171–196.

Mickler, S. E. (1993). Perceptions of vulnerability: Impact on AIDS-preventive behavior among college adolescents. *AIDS Education and Prevention, 5*(1), 43–53.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgments of numerical inequality. *Nature, 215*, 1519–1520.

Rothman, A. J., & Kiviniemi, M. T. (1999) Treating people with information: An analysis and review of approaches to communicating health risk information. *Journal of the National Cancer Institute Monographs, 25*, 44–51.

SAS Institute. (1994). *SAS/STAT user's guide.* Cary, NC: SAS Institute.

Schwarz, N. (1999). Self reports. How the questions shape the answers. *American Psychologist, 54*, 93–105.

Schwarz, W., & Ischebeck, A. (2000). Sequential effects in number comparison. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1606–1626.

Sedlmeier, P., & Gigerenzer, G. (2001). Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology: General, 130*, 380–400.

Shepard, R. N. (1981). Psychological relations and psychophysical scales: On the status of "direct" psychophysical measurement. *Journal of Mathematical Psychology, 24*, 21–57.

Shuford, E. H. (1961). Percentage estimation of proportion as a function of element type, exposure time, and task. *Journal of Experimental Psychology, 61*, 430–436.

Sokol, S. M., McCloskey, M., Cohen, N. J., & Aliminosa, D. (1991). Cognitive representations and processes in arithmetic: Interferences from the performance of brain-damaged patients. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 355–376.

Spence, I. (1990). Visual psychophysics of simple graphical elements. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 683–692.

Stevens, S. S. (1956). The direct estimation of sensory magnitudes-loudness. *American Journal of Psychology, 69*, 1–25.

Stevens, S. S. (1986). *Psychophysics: Introduction to its perceptual, neural, and social prospects*. New Brunswick, NJ: Transaction Books.

Takahashi, A., & Green, D. (1983). Numerical judgments with kanji and kana. *Neuropsychologia, 21*, 259–263.

Teigen, K. H. (1973). Number and percentage estimates in sequential tasks. *Perceptual and Motor Skills, 36,* 1035–1038.

Tversky, A., & Fox, C. R. (1995). Weighing risk and uncertainty. *Psychological Review, 102*, 269–283.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131.

van der Velde, F. W., van der Plight, J., & Hooykaas, C. (1994). Perceiving AIDS-related risk: Accuracy as a function of differences in actual risk. *Health Psychology, 13,* 25–33.

Varey, C. A., Mellers, B. A., & Birnbaum, M. H. (1990). Judgments of proportions. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 613–625.

Wanke, M., Schwarz, N., & Bless, H., (1995). The availability heuristic revisited: Experienced ease of retrieval in mundane frequency estimates. *Acta Psychologica, 89,* 83–90.

Received August 2, 2001
Revision received May 6, 2002
Accepted May 6, 2002 ■