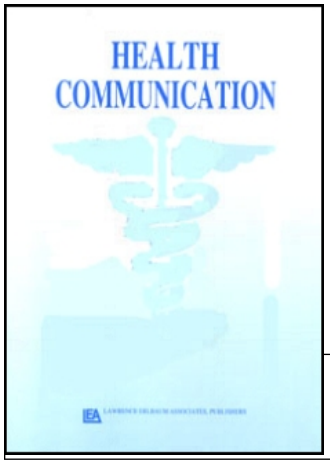

Health Communication
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t775653649

Stretched Rating Scales Cause Guided Responding

Bärbel Knäuper [a]; Christine Stich [b]; Melanie Yugo [c]; Chuck Tate [d]
[a] Department of Psychology, McGill University, Montreal
[b] Department of Psychology, Free University of Berlin,
[c] Canadian Institute for Health Information, Ottawa, Canada
[d] Department of Psychology, University of Oregon,

Online Publication Date: 01 May 2008

To cite this Article: Knäuper, Bärbel, Stich, Christine, Yugo, Melanie and Tate, Chuck (2008) 'Stretched Rating Scales Cause Guided Responding', Health Communication, 23:3, 253 — 258

To link to this article: DOI: 10.1080/10410230802056230
URL: http://dx.doi.org/10.1080/10410230802056230

Health Communication, 23: 253–258, 2008

ISSN: 1041-0236 print / 1532-7027 online
DOI: 10.1080/10410230802056230

Routledge
Taylor & Francis Group

Stretched Rating Scales Cause Guided Responding

Bärbel Knäuper
Department of Psychology
McGill University, Montreal

Christine Stich
Department of Psychology
Free University of Berlin

Melanie Yugo
Canadian Institute for Health Information
Ottawa, Canada

Chuck Tate
Department of Psychology
University of Oregon

Decision making by policymakers, public health professionals, and health care providers is often guided by the extent to which individuals feel at risk for certain adverse health events. Such health risk perceptions can be assessed in surveys using different types of probability rating scales. It has recently been suggested that rating scales that offer decomposed numeric values at the lower end of the scale (stretched scales) improve the accuracy of estimates of small risks. However, the authors suggest that respondents use the differentiated small numeric values as cues to guide them to the correct response. Study 1 supports this proposition by showing that response distributions are substantially skewed toward the lower end of stretched rating scales and have restricted variances as compared with equal-interval scales. Study 2 provides experimental evidence that scores on the stretched scale are a result of guided responding. The results show that scores on stretched rating scales are not a valid reflection of respondents' risk perceptions, but, instead, guide responses to the end of the scale that has been stretched. The findings suggest that stretched rating scales result in biased risk estimates, which may hinder effective communication about health risks between decision- and policymakers as well as between individuals and their health care providers.

One important piece of information that feeds into health decision making is the perceived risk of adverse health-related events. Individuals communicate this information to clinicians in medical settings and to decisionmakers in surveys and other forms of data collection. Obviously, it cannot be in the interest of clinicians, researchers, and policymakers to use inflated or deflated health risk estimates. Therefore, unbiased measurement approaches need to be used to elicit risk communications from individuals.

In an effort to facilitate the accurate communication of perceived health risks, researchers have developed different approaches of assessment and compared them in terms of the resulting accuracy by relating the estimates to actual statistics (e.g., Diefenbach, Weinstein, & O'Reilly, 1993; Woloshin, Schwartz, Byram, Fischhoff, & Welch, 2000). One such approach is the use of "stretched," or nonlinear, rating scales. Stretched scales offer fine-grained response options at one end of the scale (see Figure 1, for examples).

Correspondence should be addressed to Bärbel Knäuper, Department of Psychology, McGill University, 1205 Dr. Penfield Avenue, Montreal, QC, Canada, H3A 1B1. E-mail: barbel.knauper@mcgill.ca

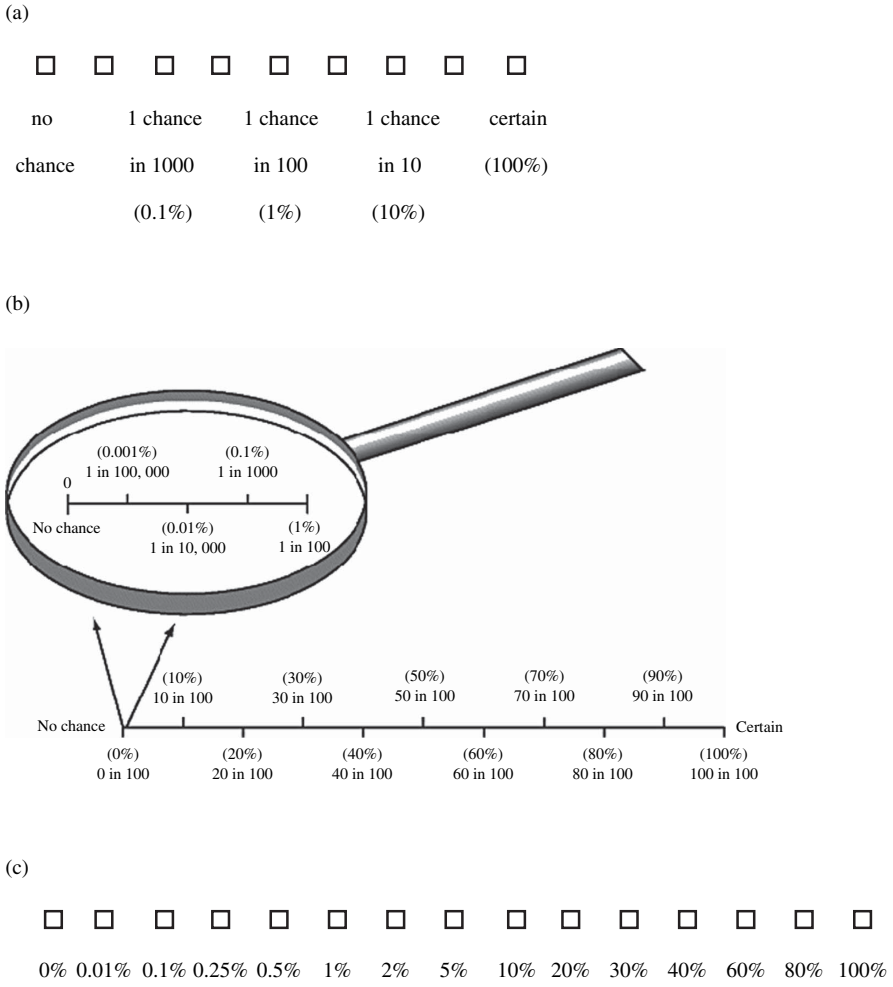

FIGURE 1 Different stretched rating scales: (a) odds scale (Diefenbach et al., 1993), (b) magnifying-glass scale (Woloshin et al., 2000), and (c) stretched scale used in these studies.

It was recently suggested that stretched scales improve the reliability and validity of estimates of small risks (e.g., Diefenbach et al., 1993; Gurmankin, Helweg-Larsen, Armstrong, Kimmel, & Volpp, 2005; Woloshin et al., 2000). This article, however, proposes that scores on stretched rating scales are not valid reflections of respondents' risk perceptions, but, instead, guide responses to whichever end of the scale has been stretched.

STRETCHED RATING SCALES

The idea behind stretched rating scales is intuitive and straightforward: When assessing the perceived risk of a rare event (e.g., the chances of having a brain tumor), only a subportion of a 0% to 100% probability scale really applies to the target event. A scale with equal intervals from 0% to 100% (e.g., with 10% intervals) does not provide respondents the opportunity to select precise small numbers (e.g., 0.001%). As a solution to this problem, researchers have proposed rating scales that offer differentiated small numbers at the lower end of the scale, thereby allowing respondents to choose among more fine-grained small numbers.

Several studies found that risk estimates for rare events are lower when assessed with stretched rating scales (e.g., Diefenbach et al., 1993; Linville, Fischer, & Fischhoff, 1993; Woloshin et al., 2000). Because it is usually found that respondents overestimate the probability of rare events (Linville et al., 1993; Quadrel, Fischhoff, & Davis, 1993; Weinstein, 1998; Windschitl & Wells, 1996), the researchers interpreted the lowering of the estimates as indicating higher judgment accuracy caused by stretching the rating scale (Woloshin et al., 2000).

We propose, however, that the highlighting of one part of the scale may create a demand characteristic to respond in this range. Responses on stretched rating scales would therefore not be valid reflections of respondents' actual risk perceptions, even though ratings might appear more accurate because the researcher who developed the scale purposefully highlighted those numeric values that match the actual statistics of the target event. This reasoning is

supported by ample evidence showing that respondents apply Grice's (1975) logic of conversation when answering questions. For example, it has been shown that respondents draw on formal features of questions such as the rating scale's graphic layout (Schwarz, Grayson, & Knäuper, 1998) as sources of information to interpret the meaning of questions (Schwarz, 1996). Applied to the differentiated low numeric values on the stretched scale, respondents may draw the "pragmatic implication" (Harris & Monaco, 1978) that the researcher probably designed the rating scale with reason and purpose and that the low numbers probably indicate the correct range of responses. Two studies test these propositions.

STUDY 1: RESPONSE DISTRIBUTIONS

Study 1 serves to examine the response distributions and variances of stretched rating scales. If responses to stretched rating scales indeed reflect guided responding, then responses on stretched rating scales should be similar to each other across participants; that is, responses should show little variability and should be skewed in the portion of the scale that has been highlighted by the researcher by the use of small numbers.

Participants, Procedure, and Measures

Samples. The first sample consists of 83 German undergraduate psychology students (mean age 24.33, male: 16.9%). The study was conceptually replicated in a second sample of 63 German undergraduate psychology students (mean age 24.87, male: 25.4%).

Procedure. Participants in Sample 1 were asked two questions about skin cancer. First, they were asked to estimate the percentage of Germans who develop skin cancer at some point in their lives. Second, they were asked to assess their perceived personal risk of skin cancer by estimating the probability that they themselves will develop skin cancer at some point in their life. Participants in Sample 2 were asked to estimate the percentage of Germans who are currently infected with HIV.

Scale versions. In both samples, one group was provided with a 15-point equal-interval rating scale with labels for 0%, 25%, 50%, 75%, and 100%, and a second group was given a 15-point stretched scale (see Figure 1c). The portion of the rating scale that corresponds with the actual statistical probabilities[1] of developing skin cancer and the percentage of Germans currently infected with HIV, that is, the lower part of the rating scale, was "stretched."

[1]Actual statistics cite the cumulative lifetime risk for skin cancer (basal cell carcinoma, squamous carcinoma of the skin and malignant melanoma) in Germany as 8% (Breitbart et al., 1997) and the prevalence of HIV infections in Germany as 0.1% (Robert Koch Institute, 2002).

Results

Responses on the stretched scale are concentrated at the lower end of the scale for all three questions (Figure 2). Replicating earlier studies, average risk estimates are lower for the stretched as compared with the equal-interval scale and thereby closer to the actual statistics: The median for the estimated lifetime risk of skin cancer is 10% ($M = 8.72$, $SD = 6.39$) when the stretched scale is used and 31.25% ($M = 31.74$, $SD = 17.73$) when the equal-interval scale is used, median test[2]: $\chi^2(1) = 23.23$, $p = .000$. Similarly, median estimates for the perceived personal risk of skin cancer are significantly smaller on the stretched scale than on the equal-interval scale, $\chi^2(1) = 31.67$, $p = .000$. For HIV, the majority of respondents using the stretched scale provide estimates between 0% and 1%. In comparison, the majority of respondents using the equal-interval scale provide estimates between 6% and 20%. As expected, the average estimates are closer to the actual statistics of 0.1%. The median is 1.0% when the stretched scale is used ($M = 2.25$, $SD = 2.12$) and 8.3% ($M = 12.18$, $SD = 6.06$) when the equal-interval scale is used, $\chi^2(1) = 48.01$, $p = .000$. Responses are thus lower, and thereby more in line with the actual statistics for the event, when the stretched scale is used.

Response distributions. Various parameters indicate that the distributions shown in Figure 2 for the stretched rating scales show little variability and are skewed (Table 1). Focusing first on the prevalence of skin cancer in Germany, the distribution for the stretched rating scale is extremely positively skewed (1.39), whereas scores on the equal-interval scale are much less skewed (0.68). Similarly, scores on the stretched rating scale have a positive (leptokurtic) Kurtosis (2.09), whereas the equal-interval scale distribution does not (0.10). Furthermore, scores on the stretched scale are more restricted in range (Moses test: $p = .000$) and have a smaller variance than scores on the equal-interval scale, $F(1, 81) = 25.33$, $p = .000$. The distribution differs significantly from a normal distribution (Kolmogorov-Smirnov $Z = 1.86$, $p < .004$), whereas this is not the case for the equal-interval scale ($Z < 1$). Results are similar for skin cancer and HIV (see Table 1). In sum, stretched scales produce significantly less variability, more skewness, and a more restricted range when compared to equal-interval scales. Study 2 examines the mechanism by which stretched scales guide responses.

STUDY 2: GUIDED RESPONDING

Study 2 examined the assumed mechanism by which stretched scales guide responses by manipulating the

[2]We performed median tests rather than parametric tests because of the highly skewed distribution of the risk measure in the stretched scale condition as compared with the equal-interval scale condition.

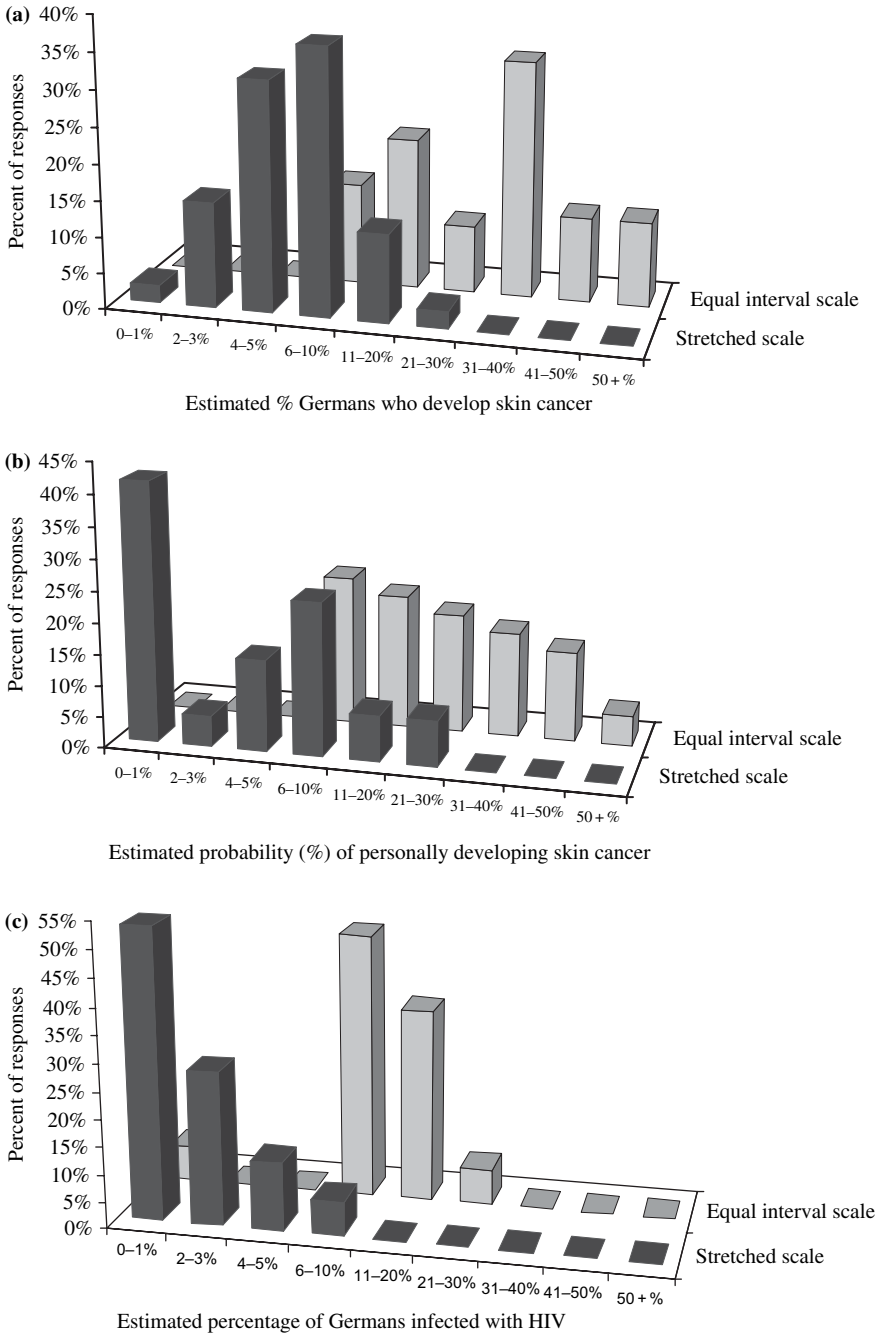

FIGURE 2 Response distributions of (a) the estimated lifetime prevalence of skin cancer of Germans, (b) the estimated personal risk of developing skin cancer, and (c) the estimated prevalence of HIV in Germany as a function of response format.

informative value of the stretched scale. In specific, Study 2 aims to demonstrate that estimates are not lowered toward the highlighted low end of the scale when the informational value of the scale is previously called into question by using the same scale for assessing common events. This would demonstrate that stretched scales guide respondents to use the highlighted end of the scale (unless its informative value is called into question).

Participants

Participants were 138 Canadian university students (mean age 21.22, male: 24.6%).

Materials and Procedure

An experimental design with one control and two experimental conditions (described following) was used. In all

TABLE 1
Descriptive Statistics of Scores in Samples 1 and 2 as a Function of Scale Version

Scale Version	*M (SD)*	*Mdn*	s^2	*Skew (SE)*	*Kurtosis (SE)*	*Range (Min/Max)*
Sample 1, Question 1[a]						
Stretched	8.72 (6.39)	10.0	40.8	1.39 (0.37)	2.09 (0.72)	29.5 (0.5/30.0)
Equal-interval	31.74 (17.7)	31.3	314.43	0.68 (0.37)	0.10 (0.72)	66.67 (8.33/75.0)
Sample 1, Question 2[b]						
Stretched	7.20 (8.57)	5.0	73.52	1.54 (0.37)	1.73 (0.72)	30.0 (0/30.0)
Equal-interval	26.09 (15.7)	25.0	245.70	0.62 (0.37)	−0.68 (0.72)	54.17 (8.33/75.0)
Sample 2[c]						
Stretched	2.25 (2.12)	1.0	4.51	1.02 (0.41)	0.39 (0.81)	7.4 (0.1/7.5)
Equal-interval	12.18 (6.06)	8.30	36.7	0.25 (0.42)	−0.21 (0.82)	24.99 (0.01/25)

[a]Estimated lifetime risk of skin cancer of Germans. [b]Estimated personal lifetime risk of skin cancer. [c]Estimated prevalence of HIV in Germany.

three conditions, respondents were first asked to estimate the prevalence of three commonly occurring health problems (i.e., percentage of Canadians experiencing a stomachache in a given year, percentage catching a cold per year, percentage experiencing at least two headaches in a given year) and subsequently estimated the prevalence of a rare disease (percentage of Canadians infected with HIV). A 20-cm line[3] was used as the basic measurement device for all four questions in all experimental conditions. Only the numeric values along the 20-cm line were varied between the three conditions (see following).

Control and Experimental Conditions

Participants were randomly assigned to either a control (baseline) condition or one of two experimental conditions that varied the informational value of the stretched rating scale. In the *baseline* condition, the common events and the rare event (HIV) were assessed using an equal-interval scale with 0%, 50%, and 100% labeled. In both experimental conditions, HIV was assessed using a stretched scale similar to that shown in Figure 1c. In the *tailored* condition, the differentiation of the numeric values was matched with the respective base rate frequency of the three common events and of HIV, respectively. That is, the incidence of the common health problems was assessed using a scale that was stretched at the upper end, whereas HIV was assessed with a scale that was stretched at the lower end. Median estimates for HIV should be closest to the actual statistics in this condition. In the *untailored* condition, both the common events and the rare event were assessed using a rating scale stretched at the lower end of the scale (i.e., a scale similar to Figure 1c was used for assessing all four events). Median estimates for HIV should be relatively high as compared with the other experimental condition.

[3]The centimeter line approach was chosen as a conservative test of the hypothesis because it provides a truly continuous scale, allowing respondents theoretically to mark any percentage response from 0% to 100%, including very small percentages in the decimals (e.g., values smaller than 1%).

Scoring

To generate percentage estimates from respondents' marks on the equal-interval scales (baseline condition), we measured the distance in millimeters from the respondents' mark to the scale's left-most anchor. Dividing this value by the total length of the scale (i.e., 20 cm) converts the measured distance into a probability measure. Probability estimates on the stretched scales (experimental conditions) were the respective numeric values underneath the respondents' mark.

Results

Results show that the median percentage estimates of the prevalence of HIV differ as a function of condition, Kruskal-Wallis $\chi^2(2) = 25.93$, $p = .000$. Specifically, the tailored condition produced the lowest median percentage estimate ($Mdn = 1.00$), the baseline condition produced the highest median ($Mdn = 10.38$), with the median for the untailored condition in between ($Mdn = 4.13$). Consequently, and as predicted, the median percentage estimates yielded by the tailored condition are closest to the Canadian prevalence rates of 0.2% (Health Canada, 2002). Also as predicted, median tests show that estimates for HIV prevalence in the tailored condition are reliably lower than in the other two conditions, that is, tailored versus untailored: $\chi^2\ (1) = 6.87$, $p < .009$; tailored versus baseline: $\chi^2\ (1) = 18.65$, $p = .000$. The estimates are also lower for the untailored condition than for the baseline condition, $\chi^2\ (1) = 11.58$, $p < .001$, suggesting that the differentiated low values on the stretched scale still had some power to lower the estimates as compared with the equal-interval scale in the baseline condition. The effect of guided responding conceptually replicates for the common health problems, with the medians being higher in the tailored condition (99.01, 99.46, and 75.43, respectively, for stomachaches, headaches, and colds) than in the untailored condition (80.74, 81.92, and 62.31). These results show that stretching rating scales at the upper end also shifts responses, this time in the direction of the highlighted range of numeric values at the upper end of the scale.

DISCUSSION

Our Study 1 data show that stretched scales produce significantly less variability, more skewness, and a more restricted range when compared to equal-interval scales. This suggested that respondents use the differentiated small numeric values as cues to guide them to the correct response. Study 2 provided experimental evidence that scores on the stretched scale are a consequence of guided responding. Results from both studies suggest that scores on stretched rating scales are not a valid reflection of a respondent's risk perceptions. If higher concordance with actual statistics is found when stretched scales are used, it does not reflect higher judgment accuracy but rather that the researcher highlighted the "correct" range of numerical values on the scale, that is, guided respondents to the correct range of responses.

This research shows that stretched scales do not provide the answer to the quandary of finding unbiased methods for assessing risk perceptions. If respondents are cued by the structure of the response scale to choose certain classes of probabilities (either high or low) over others, then the response has been guided by the measurement device, not the actual risk perceived. This has serious implications for the communication of health risks to decision and policymakers, as interventions and health promotion programs would be based on false assumptions about what people believe are risks associated with certain behaviors. Similarly, results obtained when assessing risk estimates with stretched scales invite the false assumption that individuals indeed know the probability of certain adverse health events (and that therefore the provision of knowledge is not necessary), whereas in fact they do not know and were only guided to communicate a certain risk estimate by the measure itself. In sum, these data suggest that stretched rating scales result in biased risk estimates, which may hinder effective communication about health risks between decision and policymakers as well as between individuals and their health care providers.

ACKNOWLEDGMENTS

Financial support for this study was provided, in part, by a New Opportunities Fund (Grant 4015) from the Canadian Foundation for Innovation and a grant by the Social Sciences & Humanities Research Council of Canada (SSHRC 410200209) to the first author. We thank Moon-Ho Ringo Ho, Jens Eisermann, and Norbert Schwarz for most valuable feedback on an earlier version of the manuscript and Surkhraj Cheema for her help with the data collection and analysis.

REFERENCES

Breitbart, M., Garbe, C., Buttner, P., Weiss, J., Soyer, H. P., Stocker, U., et al. (1997). Ultraviolet light exposure, pigmentary traits and the development of melanocytic naevi and cutaneous melanoma. A case-control study of the German Central Malignant Melanoma Registry. *Acta Dermatologica Venereology*, *77*, 374–278.

Diefenbach, M. A., Weinstein, N. D., & O'Reilly, J. (1993). Scales for assessing perceptions of health hazard susceptibility. *Health Education Research: Theory and Practice*, *8*, 181–192.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics*. New York: Academic.

Gurmankin, A. D., Helweg-Larsen, M., Armstrong, K., Kimmel, S. E., & Volpp, K. G. M. (2005). Comparing the standard rating scale and the magnifier scale for assessing risk perceptions. *Medical Decision Making*, *25*, 560–570.

Harris, R. J., & Monaco, G. E. (1978). Psychology of pragmatic implication: Information processing between the lines. *Journal of Experimental Psychology: General*, *107*, 1–27.

Health Canada. (2002). *HIV and AIDS in Canada: Surveillance report to June 30, 2002*. Division of HIV/AIDS Epidemiology and Surveillance, Ottawa, Ontario, Canada: Centre for Infectious Disease Prevention and Control, Population and Public Health Branch, Health Canada, November 2002.

Linville, P. W., Fischer, G. W., & Fischhoff, B. (1993). AIDS risk perceptions and decision biases. In J. B. Pryor & G. D. Reeder (Eds.), *The social psychology of HIV infection* (pp. 5–38). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Quadrel, M. J., Fischhoff, B., & Davis, W. (1993). Adolescent (in)vunerability. *American Psychologist*, *48*, 102–116.

Robert Koch Institute. (2002). HIV Infektionen und AIDS-Erkrankungen in Deutschland: Aktuelle epidemiologische Daten [HIV infections and AIDS cases in Germany: Current epidemiological data]. *Epidemiologisches Bulletin* (Sonderausgabe B/2002). Berlin, Germany: Author.

Schwarz, N. (1996). *Cognition and communication: Judgmental biases, research methods, and the logic of conversation*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Schwarz, N., Grayson, C., & Knäuper, B. (1998). Formal features of rating scales and the interpretation of meaning. *International Journal of Public Opinion Research*, *10*, 177–183.

Weinstein, N. D. (1998). Accuracy of smokers' risk perceptions. *Annals of Behavioral Medicine*, *20*, 135–140.

Windschitl, P. E., & Wells, G. L. (1996). Measuring psychological uncertainty: Verbal versus numeric methods. *Journal of Experimental Psychology: Applied*, *2*, 343–364.

Woloshin, S., Schwartz, L. M., Byram, S., Fischhoff, B., & Welch, H.G. (2000). A new scale for assessing perceptions of chance: A validation study. *Medical Decision Making*, *20*, 298–307.