Interactive graphics for communicating health risks

Jessica S. Ancker

Submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy
under the Executive Committee of the Graduate School of
Arts and Sciences

COLUMBIA UNIVERSITY
2009
UMI Number: 3374085

Copyright 2009 by
Ancker, Jessica S.

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy
submitted. Broken or indistinct print, colored or poor quality illustrations
and photographs, print bleed-through, substandard margins, and improper
alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if unauthorized
copyright material had to be removed, a note will indicate the deletion.
ABSTRACT

Interactive graphics for communicating health risks

Jessica S. Ancker

Materials for consumer informatics, patient decision support, and health promotion frequently incorporate quantitative risks such as percentages, rates, or proportions. These risks are frequently illustrated with stick figures, bar charts, or other graphics. However, risk communication models (such as the extended parallel process model) and decision models (such as prospect theory) generally focus on features of the verbal message, while failing to explore the effects of number format and illustration design. That these factors are important is shown by growing literatures that describe consumers' difficulty comprehending numbers, as well as strong effects of numeracy and risk graphic design on perceived risk and choices. This dissertation proposes an integrated model of risk communication that draws from cognitive psychology and health behavior models to consider contributions of the verbal message, features of the graphic illustration, and the audience's numeracy. Next, the dissertation presents an ontology of features of risk graphics (including animation and interactivity) and their cognitive/perceptual effects, developed through a systematic literature review. Third, a qualitative formative study was conducted that resulted in the design of several animated and interactive risk graphics with applications in Web-based communication. Finally, a questionnaire study was conducted to assess the effect of the interactive graphics on risk estimates, risk feelings, and decisions, and interactions with numeracy. Numeracy strongly affected risk estimates, risk feelings, and decisions, with lower numeracy correlated with higher perceived risks. Interaction with one of the interactive graphics affected risk perceptions and narrowed differences between high- and low-numeracy respondents. Computer-based graphical displays such as the ones developed in this project have the potential to be applied in informatics interventions for health education, tailored health and risk communication, shared medical decision-making, and patient decision support. The methods used are also promising for assessing effects of other scientific data graphics.
Table of Contents

CHAPTER 1. PROBLEM AND SIGNIFICANCE .......................................................... 1

CHAPTER 2. BACKGROUND: THEORETICAL GROUNDING FOR RESEARCH ON RISK COMMUNICATION THROUGH GRAPHICS .................................................. 3

2.1 Prospect theory: decision-making under uncertainty ........................................ 3

2.1.1 Loss-framing and screening behaviors ...................................................... 5

2.1.2 Gain-framing and disease prevention behaviors ......................................... 6

2.2 The extended parallel process model of health communication ....................... 7

2.3 Modifications to the extended parallel process model ...................................... 10

2.4 Cognitive skills affect comprehension ........................................................... 11

2.5 Perception mediates representation effects ................................................. 13

2.6 Chapter summary ....................................................................................... 18

CHAPTER 3. BACKGROUND (CONT.): MECHANISMS OF REPRESENTATION EFFECTS... 19

3.1 Salience .................................................................................................. 19

3.2 Cognitive load and demands on knowledge .................................................. 20

3.3 Visual perception and illusions .................................................................... 24

3.4 Representations and learning ...................................................................... 25

3.5 Chapter summary ....................................................................................... 27

CHAPTER 4. RESEARCH METHODS .................................................................. 28

4.1 Systematic review and ontology .................................................................... 28

4.1.1 Objective and research aims ................................................................... 28

4.1.2 Methods ................................................................................................ 28

4.2 Qualitative study ...................................................................................... 29

4.2.1 Objective and research aims ................................................................... 29

4.2.2 Methods ................................................................................................ 30

4.3 Questionnaire study .................................................................................. 34
# Table of Contents

4.3.1 Objective and research aims .................................................................................. 34

4.3.2 Methods ................................................................................................................. 36

CHAPTER 5. RISK COMMUNICATION ONTOLOGY ......................................................... 43

5.1 Introduction and scope ............................................................................................... 43

5.2 Four purposes of risk communication ....................................................................... 46

5.3 Risk descriptions ......................................................................................................... 48

5.3.1 Probability as component of risk description ......................................................... 50

5.3.2 Cognitive effects of probability formats ................................................................. 51

5.4 Types of probability numbers .................................................................................... 53

5.4.1 Cognitive effects of probability number formats ...................................................... 54

5.5 Types of graphics used in risk communication .......................................................... 59

5.5.1 Probability graphics and their properties ............................................................... 60

5.5.2 Cognitive and perceptual effects of probability graphics ....................................... 66

5.6 Summary and implications ......................................................................................... 80

CHAPTER 6. RESULTS OF QUALITATIVE STUDY ....................................................... 85

6.1 Themes ...................................................................................................................... 85

6.2 Formative development .............................................................................................. 87

CHAPTER 7. RESULTS OF QUESTIONNAIRE STUDY ................................................. 90

7.1 Test-retest reliability .................................................................................................. 90

7.2 Recruitment and demographics ............................................................................... 91

7.3 Preliminary analyses of numeracy and risk perception responses .......................... 93

7.3.1 Modification in the assessment of numeracy ......................................................... 93

7.3.2 Factor analysis of perceived risk questions ......................................................... 94

7.4 Primary outcomes: numeracy, graphics, risk perception, and intention ................ 95

7.4.1 Numeracy and risk perception .............................................................................. 95

7.4.2 Effects of graphics on risk perception ................................................................. 97

7.4.3 Intention to take preventive action ...................................................................... 98
LIST OF FIGURES

2.1 Extended parallel process model................................................................. 8
2.2 Comprehension of health information.......................................................... 15
2.3 Three representations of a Bayes' Theorem problem........................................ 16
2.4 Revised risk communication model.............................................................. 18
3.1 Non-proportional graphics........................................................................... 23
3.2 Configural plot............................................................................................. 24
3.3. Differences between curves....................................................................... 25
5.1 Examples of graphic risk descriptions............................................................ 44
5.2 Ontology of risk communication messages.................................................. 45
5.3 Coding for sample risk descriptions............................................................... 50
5.4 Types of probability descriptions in health risk communication....................... 51
5.5 Formats (categories) of numbers in quantitative probability descriptions............... 54
5.6 Non-quantitative warning graphics............................................................... 60
5.7 Ontology of warning graphics and probability graphics................................. 61
5.8 Components of graphics.............................................................................. 62
5.9 Level of measurement.................................................................................. 63
5.10 Mode.......................................................................................................... 64
5.11 Proportionality............................................................................................ 64
5.12 Part-to-whole relationship.......................................................................... 65
5.13 Scaling........................................................................................................ 65
5.14 Uncertainty.................................................................................................. 66
5.15 Scaling and proportions.............................................................................. 76
6.1 Interactive search graphic.............................................................................. 89
7.1 Order of question blocks ............................................................................. 100
7.2 Correlations between verbal risk estimates and numeric ones........................ 105
8.1 Risk graphics in context.............................................................................. 127
LIST OF TABLES

4.1 Hypergeometric probabilities of different numbers of clicks in the search graphic .......................... 39
5.1 Examples of risk descriptions using words and numbers ................................................................. 43
7.1 Item correlations for responses from 9 subjects .................................................................................. 90
7.2 Characteristics of on-line and clinic study samples .......................................................................... 92
7.3 Percentages answering correctly in three studies .............................................................................. 93
7.4 Distributions of correct responses in two studies ............................................................................ 94
7.5 Factor loadings, Story 1 ................................................................................................................... 94
7.6 Factor loadings, Story 2 ................................................................................................................... 94
7.7 Risk perceptions and intention by numeracy in all groups and within Group 4 ...................... 98
7.8 Attitudes toward the graphics (univariate analyses) ........................................................................ 104
7.9 Average inaccuracy in estimates of unlabeled graphics .................................................................. 106
7.10 Differences in estimates of random and sequential graphics of same proportion .................. 106
7.11 Overestimates of unlabeled graphics .............................................................................................. 107
8.1 Probabilities associated with different numbers of clicks in the search graphic ......................... 122
ACKNOWLEDGMENTS

No dissertation can be completed without the generosity of a surprising number of formal and informal supporters. It gives me great pleasure to acknowledge their importance to this project.

My advisor, Rita Kukafka, has been unfailingly supportive, valuing the strengths I brought to the project while also encouraging me to expand my skills and perspectives. My research benefited from perspectives from committee members David Kaufman, Stephen Johnson, and Elke Weber at Columbia, and Bradford Hesse at the National Cancer Institute. I am especially grateful to Dr. Weber, who contributed not only the inspiration for one of the original experiments but also rigorous methodological brainstorming sessions and the resources of her Virtual Laboratory at the Columbia Center for the Decision Sciences. My thoughts on graphics were formed in part by a collaboration with Justin Starren. Programming could never have been completed without the cheerful assistance of Jianhua Li and the willingness of Yalini Senathirajah to answer a nearly infinite number of questions. Connie Chan helped me tremendously with the qualitative study, and Paul Teixeira helped me clarify my ideas on the questionnaire studies. Alwyn Cohall, co-director of the Harlem Health Promotion Center, allowed me to conduct formative studies at the center, where Nisha Beharie served as an outstanding focus group moderator. Martin Pusic generously facilitated my recruiting patients from the waiting area of Columbia’s Department of Pediatric Emergency Medicine, and Ronnie Myers did the same at the clinics of the College of Dental Medicine. The graphics study was supported by the Agency for Healthcare Research and Quality (R03-HS016333; PI, Rita Kukafka). My doctoral training was supported by the National Library of Medicine (NLM training grant T15-LM0070790). I am also grateful to the Robert Wood Johnson Foundation for contributing to the NLM training program in public health and sponsoring additional training seminars for the NLM public health fellows.

My mentor and former employer, Tom Lang, set me on this path by introducing me to concepts and methods in measuring the efficacy of written communication, as well as to the initially counterintuitive idea that statistics can be fun.

Finally, I never would have had the courage to start graduate school at all, or stick it through to the end, without the love and support of my husband, John Affleck.
CHAPTER 1. PROBLEM AND SIGNIFICANCE

Communicating about risks is an important part of health promotion, decision support, informed consent, and other health communication activities. Perceived risk is a potential motivator of health behavior change and decision making in many health models (1-3), and a recent meta-analysis confirmed that perceived risk has a strong relationship with behavior (4). Patients should be able to understand and compare risks when making decisions such as choosing between treatments (5), agreeing to screening tests (6), understanding insurance alternatives or healthcare quality indicators (7), and granting informed consent for treatment or research (8).

Because of the importance of perceived risk, tools for consumer informatics and patient decision support frequently display quantitative risk information (probabilities, rates, or proportions). One example is a personalized heart attack risk calculator based on the National Cholesterol Education Guidelines (9) (http://hp2010.nhlbihin.net/atpiii/calculator.asp). The risk is shown first as a percentage (e.g., “20%”), then as a rate (“20 of 100 people with this level of risk will have a heart attack in the next 10 years”).

However, placing this information in a risk communication message does not ensure that recipients will understand it or act upon it. The effectiveness of risk communication is influenced by features of the verbal message, including the relative weight of concepts such as susceptibility to a threat and efficacy of the remedy for that threat (3, 10). The importance of verbal message features has been extensively explored in the literature on framing and message features (11-14).

Audience competencies such as literacy and numeracy can also create barriers to risk communication (15, 16). Skills such as ability to calculate and manipulate percentages vary widely among the public (17-19). Numeracy and educational level are associated with comfort with (20, 21) and understanding of (18, 22) probability and percentages.

Although graphics can help illustrate these concepts (5, 23, 24), poor graphic design can end up impairing comprehension (25), inflating risk perception (26), or altering choices (27-30). For example, survival curves are excellent for illustrating changes in risk over time for expert audiences, but many consumers cannot understand them (25).

Although features of the verbal message, features of the illustration, and cognitive
competencies are all known to affect risk communication, they have not been integrated into a single model because the relevant literature derives from different disciplines. This dissertation proposes an integrated model of risk communication that combines these factors (Chapter 2).

In addition, the dissertation presents a systematic review of the literature on cognitive and perceptual effects of health risk graphics (Chapter 5). The results are presented as an ontology of risk graphic features. An ontology is a rigorous classification of the objects and concepts in a particular domain, as well as the relationships between them. This ontology serves as a map of the conceptual space of risk graphics and suggests hypotheses about the likely effects of various design choices on risk perception and decision-making.

Some of the features included in the ontology are relatively new ones available only with computer-based graphics: animation and interactivity. Educational literature suggests that these features may help engage consumers, and decision science findings (31, 32) suggest that they may influence risk perception and risk-related decision making. The integration of animation and interactivity into risk graphics is explored in this dissertation in a formative qualitative study (Chapter 6). The goal was to develop prototypes for interactive risk communication modules that might be appropriate for Web-based health promotion. The qualitative assessment explored consumer attitudes toward and interpretations of different graphics, as well as usability.

Finally, the dissertation presents a questionnaire study (Chapter 7) to measure the effects of these novel graphics. Participants were asked to make decisions while viewing either interactive or static (control) graphics. The primary outcome variables were perceived risk and decision, with numeracy and health literacy among the independent variables. Two samples were recruited: one on line, and a second from NewYork-Presbyterian Hospital clinic waiting rooms.

Informatics interventions that exploit these findings to promote the clear communication of the probabilities of health outcomes would enable patients to use quantitative information productively as part of health-related decision-making. Animation and interactivity are powerful computer tools that could be applied to this problem.
CHAPTER 2. BACKGROUND: THEORETICAL GROUNDING FOR RESEARCH ON RISK COMMUNICATION THROUGH GRAPHICS

This project focuses on the effects of visual and interactive communication on health-related decision-making. Because it spans several disciplines, no single theoretical model covers the relevant territory. In this chapter, I will establish the theoretical grounding for this research by discussing concepts and constructs from theories of decision-making, health communication, and cognitive skills.

Two well-validated models form the basis for this discussion. The first is prospect theory (11), an empirically derived model of decision-making under conditions of risk or uncertainty. This model comes from the research domain of decision science. The second is the extended parallel process model of health communication (3, 10), which explains and predicts the effects of informational messages about health risks on health protective behavior. The EPPM originates from the field of behavior change but is primarily used to describe one-time "go/no-go" decisions (such as the decision about whether to get a vaccine) rather than long-term sustained behavior change.

In this section, I will demonstrate that these models have areas of similarity and have relevance for certain types of health decisions. However, I will also argue that they both have the same limitation: neither fully considers cognitive factors (literacy and numeracy) and perceptual factors (the representation effect) that have been shown to mediate comprehension of written and numeric information. A revised model is presented that integrates these factors.

2.1 Prospect theory: decision-making under uncertainty

Prospect theory is the classic model of decision-making under uncertainty (11). Prospect theory demonstrates that decision-makers who are given the choice between two uncertain options integrate the probability and the magnitude of possible outcomes into a "utility" or "perceived utility" for that option. Kahneman and Tversky showed that the probability of each outcome is transformed into perceived probability through a nonlinear weighting function called \( w \). So, for example, very small probabilities are often inflated into larger perceived probabilities. In addition, the value of an outcome (e.g., the monetary gain or loss) is transformed into perceived value
through a second nonlinear weighting function \( v \). For each of the \( k \) possible outcomes of Option 1, perceived probabilities and perceived values are multiplied, then summed, to produce a utility.

**Equation 2.1**

\[
\text{Utility of Option 1} = w_1 \times v_1 (\text{probability of outcome 1}_1) \times v_1 (\text{value of outcome 1}_1) + \\
+ w_2 \times v_1 (\text{probability of outcome 1}_2) \times v_1 (\text{value of outcome 1}_2) + \ldots + \\
+ w_k \times v_1 (\text{probability of outcome 1}_k) \times v_1 (\text{value of outcome 1}_k)
\]

The utility of Option 2 is produced in the same way on the basis of its possible outcomes. The decision is then based upon the comparison of the two utilities. The decision-maker will choose the option with the higher utility, or, if the utilities are equal, will be indifferent.

Prospect theory has been defined and explored through a research paradigm called the standard gamble, in which decision-makers are presented with choices between economic gambles (such as "a 10% chance of winning $100," compared with "a 20% chance of winning $50"). However, prospect theory or elements of it have been applied to decisions in a variety of domains including health (13, 14, 33-35).

In particular, prospect theory's description of the framing effect has been extremely influential in health communication. The framing effect is comprised of two components. First, a perceived loss of a certain size is more influential on decision-making than a perceived gain of the same size, an effect called *loss aversion* or the "losses loom larger" effect. Second, the perception of whether something is a loss or a gain from the current condition can often be altered through language or other manipulations, an effect called *reference point shift*. A simple manipulation might be to give research participants a souvenir mug. Those who have been given the mug will tend to ask for more money to sell it than they would pay to buy the mug in the first place, showing that the potential loss of a mug belonging to them feels greater than the potential gain of the same mug (the "endowment effect"). Language is also effective at creating framing manipulations. A medical procedure may feel more threatening when framed negatively as having a 10% mortality rate than it does when it is framed positively as having a 90% survival rate.

One classic health-related example describes a pair of hypothetical public health interventions among a population of 600 people in terms of either the lives lost (the loss frame) or lives gained (the gain frame) (36). In the loss frame, Intervention A gives a 33% chance of no
deaths and a 66% chance of 600 deaths; Intervention B carries a certainty of 400 deaths. The two options are normatively equivalent because they each result an average expectation of 400 deaths. However, in the loss frame, Intervention A seems more attractive – a certain loss seems worse than the chance of a bigger loss.

In the gain frame, Intervention A carries a 33% chance of saving all 600 patients and a 66% chance of saving no one, while Intervention B will save 200 patients for sure. In the gain frame, it is Intervention B that appears more attractive. The certain gain seems more attractive than the chance to make even greater gains. In other words, people are risk-seeking when weighing information about losses but risk-averse when weighing information about gains (36).

2.1.1 Loss-framing and screening behaviors

A recent systematic review shows that loss-framed messages have been effective in promoting real-world screening or disease-detection behaviors such as mammography and breast self-exam, although generally with modest effect sizes (34). In one study, women who were not getting the recommended number of mammograms for their age were recruited through their workplace to participate in a study. They were randomly assigned to watch a loss-framed mammography video (e.g., “failing to detect breast cancer early can cost you your life”) or a gain-framed one (e.g., “detecting breast cancer early can save your life”) (13). Among the 133 women in the study, 66% of those who watched the loss-framed video and 52% of those who received the gain-framed one got mammograms within a year (13). This effect size is moderate but still fairly impressive given that it was the result of a single viewing of a video, with results measured 1 year later. Similarly, Meyerowitz and Chaiken increased frequency of breast self-exams among undergraduate women by using a loss-framed informational booklet (57% reported an increase in frequency) compared to gain-framed, neutral, or no booklets (38%, 39%, and 29% respectively) (14). However, this study was limited by its reliance on self-report of frequency of breast exams.

Two possible mechanisms for this effect are consistent with prospect theory. The first is loss aversion: avoiding a loss is more motivational than trying to achieve a gain of similar magnitude.
The second is risk aversion and reference point shift. The origin of this argument is the finding that many people view disease-detection behaviors as inherently risky (13, 14, 37). For example, surveys suggest that many women avoid doing breast exams because they fear finding a lump (14). In particular, in one study, the survey item “I don’t want to take the risk of finding a lump” correlated fairly strongly (r = 0.64) with low frequency of breast self-exam (14). Thus, the investigators suggest, performing breast exams feels like risk-seeking behavior because of the chance of finding a lump. In the gain frame, women assume that they are in good health, so not performing a breast exam feels like the sure option: it carries zero chance of detecting cancer.

Loss-framing the breast exam materials shifts the reference point, according to this argument. Rather than optimistically assuming their own good health, women now feel that they may have undetected cancer today. That is, in the loss frame, women feel that they are choosing between the potential loss of finding a lump on the breast exam today (option with an uncertain loss), and the perceived certainty that doing nothing means failing to detect breast cancer (option with a certain loss). In this loss frame, risk-seeking dominates, motivating women to perform breast exams. One goal may be to try to return to their previous reference point of perceived certainty about their own health (14).

Others have posited that the same mechanisms extend to other screening or secondary prevention activities such as mammography, Pap smears, and colorectal cancer screening (37, 38). A systematic review found 6 studies in which loss-framing modestly improved uptake of real-world disease-detection behaviors (34).

2.1.2 Gain-framing and disease prevention behaviors

By contrast, the systematic review suggests that gain-framing is more successful in promoting primary prevention or health-promotion behaviors. For example, Detweiler et al found this effect in an unusual field study in which 217 beachgoers at a public beach were given booklets with either gain-framed or loss-framed information about using sunscreen, along with coupons for free sunscreen (38). Gain-framed messages described either how sunscreen increases the chances of desirable health and beauty effects or how it reduces the chances of
undesirable effects. Loss-framed messages described either how skipping sunscreen increased the chances of bad outcomes or how it reduced the chances of staying healthy. In the results, 71% of people in the gain-framed group walked to a table at the same beach later the same day to redeem their coupons, compared to 53% of those who received loss-framed information (38). There were no substantive differences between the two types of gain-framed messages, nor between effects in men and in women. Overall, the effect was strongest among those who indicated that they had not planned to use sunscreen when they came to the beach (not surprisingly because of the likelihood of ceiling effects). The study is limited, however, by the inability to measure actual use of the sunscreen.

A loss-aversion explanation would have predicted that the loss-framed messages would have larger effects. Therefore, Detweiler et al argue in favor of the risk-aversion explanation (38). That is, since a prevention behavior such as applying sunscreen carries no chance of bad consequences (in the short term), it is perceived as the risk-averse option and is preferred in gain-frame situations. Unfortunately, this study included no neutral control group (e.g., no pamphlet, or one containing non-framed information) to establish the base rate at which people would redeem the sunscreen coupons. It is therefore unclear whether loss-framing was simply ineffective or whether it actually reduced interest in using sunscreen.

2.2 The extended parallel process model of health communication

Although the researchers in the framing studies cited prospect theory as their inspiration, they also measured constructs such as self-efficacy, perceived threat, and fear or emotional reaction to the message. These constructs are found in health behavior models such as the health belief model (1), the precaution adoption process model (2), protection motivation theory (39), and the C-SHIP model (40). This suggests that prospect theory in itself does not contain all the constructs needed to explain real-world health behavior in response to persuasive communication. In particular, utility theories focus on cognitive responses and do not model emotional responses such as fear linked to intentions and behaviors in health and other domains (33, 41-43).
The extended parallel process model was developed specifically to predict responses to communications (such as public service announcements) about protective actions against health hazards, when these communications involve fear appeals (3, 10, 41, 44). EPPM draws psychological/behavioral constructs from other behavior models, particularly protection motivation theory (39, 45), and in addition it also defines the relevant characteristics of the messages. EPPM identifies two decision processes, one primarily analytic/cognitive (top process in Figure 2.1) and one primarily emotional (bottom process in Figure 2.1), which can be differentially triggered depending on the characteristics of the message and message processing. This model advanced previous models by allowing predictions about which fear appeals are likely to fail or backfire, thus helping to resolve contradictions in the literature on fear appeals and persuasion (Figure 2.1) (3, 10, 44).

![Diagram of EPPM](image)

**Figure 2.1 Extended parallel process model.** The recipient's "first appraisal" is of threat; a small threat leads to no action (bottom). If threat seems large, a "second appraisal" of efficacy takes place. If efficacy appears sufficient, fear is reduced and the message is accepted. If efficacy seems insufficient, recipients are 'defensively motivated' to reduce fear by rejecting the message. Illustration from (46).

In EPPM, health promotion messages contain two relevant components (left side of Figure 2.1): information about the health hazard, and information about a particular recommended response to reduce that hazard. For example, a message might describe breast cancer as the health hazard, and it might describe mammography as the recommended action. The hazard information consists of information about its severity (e.g., a statement that breast cancer can be fatal) and about the consumer's susceptibility to that hazard (e.g., a description of the chances of getting it). These two pieces of information are integrated into a mental construct called *perceived*
threat. Information about the recommended response contains an element called response efficiency (e.g., whether mammography will reduce risk of dying). It may also address the listener's self-efficacy (e.g., her ability to get or pay for a mammogram). Response efficacy and self-efficacy are processed into the construct perceived efficacy.

A first appraisal focuses on perceived threat. If it appears minimal, the message is ignored and there is no further action. If the hazard is judged to be serious or relevant, the listener will appraise the efficacy of the response described in the message. If the perceived efficacy of the response outweighs the perceived threat, the listener will tend to adopt the recommended protective action.

However, if the perceived threat outweighs the perceived efficacy, either because the threat is too high or because the efficacy is too low, the listener will tend to respond with an increased feeling of fear. Relatively rapid, automatic, and emotional processes will be engaged, prompting actions to reduce the fear (center right of Figure 2.1). The listener may minimize the threat, engage defensive avoidance mechanisms by ignoring the message, or even engage in "boomerang" responses such as increasing the frequency of the non-recommended behavior (such as smoking). In these situations, fear-based appeals backfire (10). Witte argues that deliberately adopting a health precaution to reduce the threat is primarily a cognitive or analytic response to the message, while message rejection or defensive avoidance to reduce feelings of fear is a rapid, automatic emotional response.

The threat and response components of a health message may not always be explicitly described. Listeners may, for example, draw upon internal knowledge for their information about self-efficacy. Witte has shown that the ineffectiveness of certain fear appeal campaigns can be traced to imbalances between information about threat and efficacy. For example, a message is unlikely to produce action if it does not provide enough information about response efficacy. Thus, an anti-smoking message might fail if it does not specify how quitting smoking would avert the threat of cancer, or convince smokers that it is not too late to benefit from quitting. Similarly, the message might fail if it does not boost listeners' self-efficacy, for example, concerns that they can't quit. Finally, a message is likely to fail if it overstimulates fear without also working to
increase the perceived efficacy of the response (for example, anti-smoking campaigns that feature pictures of people disfigured by smoking-related cancers) (10).

2.3 Modifications to the extended parallel process model

Prospect theory is a model of choice between two uncertain options. By contrast, EPPM is a model of persuasion in which the goal is to convince people to take a particular course of action. Information about the threat and the efficacy is integrated and informs several possible behavioral options. In particular, EPPM posits two decision processes, one primarily cognitive and one primarily emotional, which can lead to different behavioral choices. Nevertheless, the two models have some similarities, and both have been applied to “go/no-go” health decisions such as whether or not to get mammography. First, both models involve options with uncertain outcomes. For example, a message promoting mammography will often acknowledge that mammography reduces the risk of cancer-related death but will not prevent it with certainty. Second, the models contain some similar constructs. Witte suggests that her “perceived severity” is similar to the perceived (dis)utility of the health hazard (41). She also suggests that the perceived probability of the health hazard is similar to “susceptibility,” and the lower perceived probability of the hazard if the protective action is adopted is similar to “response efficacy” (41). These points of similarity suggest that some findings from the prospect theory studies cited earlier could result in useful examination of specific elements of EPPM.

For example, prospect theory suggests that costs or harms associated with recommended health behaviors may affect perceived efficacy of the health behavior. As described above, the potential short-term negative consequences of the recommended behavior – i.e., finding a disease – may be perceived as costs of the screening and may themselves elicit fear or anxiety. Mammography and breast self-exam are perceived to have immediate risks that elicit fear (14), which if large enough can lead to rejection of the message; applying sunscreen has few or no such risks. A recent survey study found that parental belief that childhood vaccines are unsafe is

---

1EPPM suggests an explanation for the widespread mockery of the U.S. Department of Homeland Security’s national threat advisory system, the color-coded terrorism alert scale. The threat is not accompanied by any precaution message, so the only available response to manage fear is to reject the message.
associated with belief that vaccines are also not as effective as advertised (47). Paying attention to costs or barriers to actions would also be consistent with the health belief model (1).

The prospect theory studies also show that efficacy of a response can be described in terms of probability of averting the harm, the amount of harm averted, or both. An example with uncertain probability of benefit is the influenza immunization, which, as patients are routinely told, does not prevent flu 100% of the time (e.g., http://www.cdc.gov/flu/about/qa/vaccineeffect.htm). Other health-related actions in which patients are routinely told that there is an uncertain chance of averting disease or death are mammography and genetic testing. Examples of modest benefit size include pain-relievers that reduce pain without eliminating it. Thus, in some real-world situations, response efficacy is expressed as probability of response, or size of response, or both.

A complete discussion about how framing might affect the EPPM constructs is beyond the scope of the current research. However, it seems possible that framing manipulations might affect the EPPM constructs perceived efficacy, perceived threat, or perceived costs of the response. The Banks study was unable to identify the mediators of the framing effect, but the researchers did not measure either perceived efficacy of mammography nor the perceived costs or dangers of mammography (13). They did find in the immediate post-video assessment that the loss frame was associated with small increases in perceived risk of breast cancer and self-efficacy about mammography. Although the researchers dismissed the changes as statistically non-significant, it is possible that small differences at baseline led to stronger intentions or memory by 12 months.

2.4 Cognitive skills affect comprehension

EPPM and prospect theory both assume that a listener or reader will comprehend the information provided in the message. However, basic cognitive skills vary among readers and affect comprehension. Any model of health communication must account for these skills.

To comprehend medical or health information, readers need several sets of skills including literacy or health literacy, and numeracy. Literacy is the skill set needed to use printed and electronic health information in simple text, oral instructions, or complex document formats such as tables and forms (48). Health literacy is more specifically the set of skills needed to obtain and
use such information to make health decisions and guide health-related behaviors. Health literacy requires not only traditional literacy skills such as reading text and navigating documents, but also a basic level of health-specific domain knowledge such as understanding of medical vocabulary (15, 16, 49-51). Health numeracy is the set of skills needed to use quantitative information in health contexts (17, 52). Both national surveys and targeted studies have identified widespread deficits in literacy, health literacy, and numeracy (18, 19, 48, 49, 53-56). Limited health literacy and numeracy have been linked with poor health outcomes and behaviors, impaired ability to manage chronic diseases, and higher hospitalization rates (57-60).

One dimension of health numeracy that is particularly relevant to risk communication is the ability to interpret percentages and probabilities (18). Even a well-educated sample displayed a wide range of performance levels on a numeracy questionnaire that focused on probabilities, suggesting that this aspect of numeracy does not correlate well with educational level (19). Up to 20% were unable to correctly answer questions such as “Which represents the larger risk: 1%, 5%, or 10%?” (19). Another relevant dimension of health numeracy is graphical literacy, or the ability to interpret graphs (17). For example, lay people often cannot answer questions about survival graphs, which are commonly used for communicating with physicians (25, 61, 62).

Numeracy skills are associated with comfort with (20) and understanding of (18, 21) quantitative risk information. For example, in a sample of women drawn at random from a database of veterans, low numeracy was associated with impaired comprehension of information about the potential risk reduction offered by screening mammography (18). The less numerate may trust verbal risk statements such as “low risk” instead of numeric ones such as “5” (20).

A potential explanation might be that people with poor numeracy skills have difficulty making emotional judgments about numbers. In a standard gamble study (55), participants were asked not only to pick the options with the best payoffs, but also to rate the attractiveness of each option. Participants with lower numeracy were more uncertain about the attractiveness of all the options (55). Emotional responses are known to be extremely important in decision-making (63), even relatively analytic decisions (64). In addition, the emotional feeling of risk may be more closely related to the mental construct of perceived risk than the perception of the quantitative
magnitude of a risk (42, 43). Feeling of risk predicts subsequent influenza vaccination better than that person’s perception of the risk magnitude (65). Thus, impaired ability to make emotional judgments from numbers might make it particularly difficult to make decisions about risk.

Another cognitive factor that is known to affect comprehension of risk-related information is depth of cognitive processing. Active learning engages central cognitive processing mechanisms that are deeper than the peripheral processing mechanisms involved in the passive receipt of information (66). In the elaboration likelihood model, Petty and Wegener focused primarily on the effect of deep processing on attitude, and showed that attitude change derived from active information processing is more stable and persistent than changes derived from passive learning. However, deep central processing has cognitive effects as well as attitude-related ones. Natter and Berry provided participants with numerical risks, then invited them to draw a graph to illustrate the risk or to answer a reflective essay question encouraging thoughtful interpretation of the risk. Such active processing improved both comprehension of and satisfaction with the information (67). In addition, studies such as these show that deep cognitive processing can be induced by the researcher.

This evidence about the effects of literacy, health literacy, numeracy, and depth of processing suggests that no model meant for applied research can afford to ignore these factors. It also suggests that some of the “individual differences” mentioned in EPPM (in the lower middle panel of Figure 2.1) are in fact the cognitive skills numeracy and health literacy.

2.5 Perception mediates representation effects

However, the reader’s skills are not the only factors determining whether the reader will comprehend medical information. A deep body of research from disciplines such as human factors and cognitive psychology shows that the format in which information is represented has a strong effect on comprehension and problem-solving (68). This phenomenon is known as the representation effect. In a simple example, time can be represented by the hands of a traditional analog clock or on a digital clock face, which are logically equivalent but cognitively different formats (69). The digital format makes it easy to determine the precise time to the nearest second
(or even hundredth of a second). However, determining the duration of an event is more difficult because it requires a series of relatively effortful cognitive steps involving arithmetic and the application of the learned convention that an hour comprises 60 minutes. (For example, the duration of a phone call that lasted from 2:53 p.m. to 3:10 p.m. could be determined by subtracting 53 minutes from 60 minutes to get 7 minutes, subtracting 0 from 10 to get 10 minutes, and then adding 10 and 7.) By contrast, analog clocks make it easy to estimate duration by representing duration as slices of a pie: one quarter, one third, and one half of an hour are particularly easy to visualize. However, a quick glance at an analog clock shows only an approximate time. Only by closely examining the position of the hands is it possible to determine time to the nearest second, and determining time to the nearest tenth of a second is impossible.

The representation effect has been recognized as relevant to health literacy, albeit not by that name. An IOM report (50) criticized the American Medical Association definition of health literacy (51) for focusing too narrowly on the role of the patient's skills without recognizing the importance of health professionals' communication skills and clear writing and design of written information. Health literacy expert David Baker, a developer of the TOFHLA health literacy test (70, 71), has made a similar argument about health literacy (72). I have similarly argued that discussions of quantitative information in health should be expanded beyond a focus on the skills of the patient (17). Although inability to use quantitative information may stem from poor skills, it may also be the result of mismatch between the patient's skills and the provider's communication skills or the artifact's information design (17). Decisions and comprehension of health information can be improved by designing information systems and artifacts to be easy to use and understand.
One simple and compelling demonstration of the representation effect in numeracy was a simple questionnaire study among women at an ob-gyn clinic (22). Only 56% could identify the higher risk when the question was phrased as “Which rate is higher, 1 in 384 women, or 1 in 112 women?” However, 73% could answer correctly when asked the numerically equivalent question, “Which rate is higher, 2.6 per 1000 women, or 8.9 per 1000 women?” The second representation made it easier for women of all educational levels to make the comparison. In addition, this study shows a strong interaction between the representation effect and education level. Among women with 9 years of education or less, only 30% answered the first question correctly and 50% answered the second one correctly (22). (The subsequent chapter of this dissertation will discuss some potential cognitive mechanisms underlying this and other representation effects.)

Another demonstration of the representation effect was an experiment conducted by Hoffrage and Gigerenzer among students and faculty at Harvard Medical School (73, 74). In this experiment, the participants were asked to solve a basic reasoning problem about diagnostic test results: given the sensitivity and specificity of a test, and the prevalence of disease, what is the probability that a patient who tests positive really has the disease? This type of problem is considered a basic element of clinical decision making (75). When the problem was presented as a set of probabilities, only 1 out of 24 participants could solve it. However, when the problem was
given in "natural frequency" format, 16 out of 24 could solve it (Figure 2.3). The researchers have described their method as "improving reasoning without instruction" (76).

**Statistical representation**

\[
PPV = \frac{\text{sensitivity} \times \text{prevalence}}{\text{sensitivity} \times \text{prevalence} + (1-\text{spec})(1-\text{prev})}
\]

\[
= \frac{(0.95)(0.002)}{(0.95)(0.002) + (0.10)(0.998)}
\]

\[
= 0.02 = 2\% \text{ chance that the woman with a positive mammogram has cancer}
\]

**Natural frequencies representation: verbal**

Of every 10,000 American women, 20 have HIV and 9980 do not. If the 20 are given a screening test, 19 will test positive. If the 9980 are given a screening test, 998 will falsely test positive. The chance that a woman who tests positive is HIV-positive:

\[
\frac{19}{19 + 998} = 0.02 = 2\%
\]

**Natural frequencies representation: graphical**

![Diagram](image)

Figure 2.3. Three representations of a Bayes' Theorem problem. Statistical presentation (top left), natural frequencies version (top right), and graphical version (bottom). PPV = positive predictive value.

Graphics can induce strong representation effects (23, 24). Well-designed graphs can improve reasoning about quantitative information for both expert and lay audiences (77-80). Different graphic representations of the same quantity can lead to very different patterns of decision-making, as well as differences in comprehension (24, 27-29). In the Stone studies, participants were shown different types of graphics portraying the safety risk attached to particular consumer products, and then were asked how much they would be willing to pay to reduce the risk. One graphic was a set of small stick figures representing those affected by the hazard; those unaffected were not portrayed in the illustration, so the part-to-whole relationship was not portrayed. This graphic was associated with a much higher willingness to pay than any graphic that did include the part-to-whole relationship. Other studies have also found representation effects with icons, dots, or stick figures for demonstrating probabilities to both lay
audiences (24, 81-84) and expert ones (30). The Elting study revealed very strong representation effects on a physician audience. The doctors were shown information from a fictitious clinical trial in which one group of patients had much worse outcomes than the other. They were asked whether the trial should be halted. The format in which the information strongly affected their decisions, with few doctors considering the risk serious enough to halt the trial with a pie chart, and many halting the trial when the same data were shown as icons (30). In a study with lay people, a hypothetical treatment decision was described through anecdotes about other patients or with both anecdotes and an icon grid showing the risks. The addition of the icon grids reduced the influence of the anecdotal information on decisions (82). Icon grids were not effective at reducing people's tendency to overreact very small probabilities (85). Such matrices appear to acceptable and even preferred in studies with the general public (81, 86), although most of the physicians in the Elting study did not like the stick figures (30).

In computer-mediated education and communication, graphics can be animated. Photographs can be replaced with movies, and drawings can be replaced with animated cartoons. Interactive animations have been used in a number of health educational materials, such as computer games and websites for self-management of diabetes (87-89) and asthma (90-94). Animation has the potential to induce representation effects. For example, Tversky has shown that animated or video instructions are sometimes more difficult to follow than static illustration of instructions (95). In a set of psychological studies, the risks and benefits in standard gamble experiments were represented by decks of cards, not by numeric or graphical descriptions of the risks (31, 32, 96). Participants learned about the riskiness of a real or virtual deck of cards by drawing cards from it. Each deck contained different proportions of winning and losing cards. The classical prospect theory studies had long shown that people overreact to small risks when told about them (suggesting that perceived probability was smaller than the probability), but in these experiential learning studies, people tended to underreact to the same risks. There were also several other reproducible differences from the decision-making patterns in the classical studies (31, 32, 96). Similar findings were recently reported by an unrelated group of
cognitive researchers (97). These differences between behavior patterns with static descriptions and interactive animated graphics are examples of representation effects linked to animation.

Figure 2.4: Revised risk communication model. Design of graphics and format of numbers are important message components, and cognitive and visual processing precedes message processing and outcomes. (Added components in yellow.)

2.6 Chapter summary

Communication with the public about health risks has been studied from a variety of theoretical perspectives. The extended parallel process model focuses on how persuasive communication affects the choice of whether to adopt a recommended preventive action in the face of a health hazard. Prospect theory describes the choice between two options with uncertain outcomes, but has also been applied to persuasion in health. In particular, the framing effect, first described in prospect theory, has become influential in the health communication literature on message features. However, neither model explicitly considers cognitive and perceptual factors such as literacy and numeracy that have been shown to mediate comprehension of written and numeric information. In fact, the extended parallel process model simply lists “individual differences” as contributors to the interpretation of risk information. Also, neither the EPPM nor prospect theory accounts for how different representations of information can affect decisions and behaviors. In the modified model described here, these factors are integrated to provide justification for the experimental sections of this dissertation.
CHAPTER 3. BACKGROUND (CONT.): MECHANISMS OF REPRESENTATION EFFECTS

The previous chapter discussed theoretical models of risk communication that justify research into the way that health risks are represented (for example, in words, numbers, or static or animated graphics). As discussed, different representations of numbers and quantities strongly affect comprehension, problem-solving, and decision-making, in a set of phenomena known as representation effects.

This section will cover some of the perceptual and cognitive mechanisms through which representations induce these effects, including salience, cognitive load, and visual perception. A final section will discuss the role of different representations in several theories of learning.

3.1 Salience

Several researchers have proposed that one mechanism is salience, the representation’s ability to draw attention to certain elements, quantities or relationships, and to minimize attention to others. An example of the salience explanation is found in a series of studies by Stone et al (27-29). Participants were shown different types of graphics portraying the safety risk associated with particular consumer products, and then were asked how much they would be willing to pay to reduce the risk. One graphic was a set of small stick figures representing those affected by the hazard; those unaffected were not portrayed in the illustration, so the part-to-whole relationship was not portrayed. This graphic produced stronger risk-reducing behavior (larger willingness to pay) than any graphic that did show the part-to-whole relationship. Stone et al attribute this result to the graphic’s ability to make the risks salient, in what they have called the “foreground effect” (27-29). Jarvenpaa has shown that in consumer graphics, viewers tend to select information on the basis of the visual salience of the graphic element, not other factors such as the importance of the information (98).

Salience effects can be particularly strong in animations because movement tends to attract visual attention. Animations that draw viewers’ attention to important information thus have the potential to improve comprehension. However, when the wrong information is highly visually salient, animation may interfere with comprehension. In studies by Lowe, the most visually salient
movement in weather maps attracted attention, but this effect sometimes detracted from comprehension when the salient information was not also the most important (99, 100). Tversky has also shown that animations and videos sometimes include too many irrelevant details, obscuring the salience of important information (95).

There is no guarantee that animation alone will call attention to the right types of information to improve comprehension or learning. In one study, self-paced educational modules of medical information about the heart with and without animated graphics produced similar knowledge scores (101). Consumers learned no more from Flash-animated patient education brochures than from printed versions (102, 103). However, in these latter studies, the consumers preferred the animated versions, suggesting that animation may play a useful role in making information more attractive to a target audience.

Salience manipulations could help novices understand a new concept or domain. Studies of expert-novice differences show that newcomers to a field of knowledge typically have difficulty sorting through information to distinguish important details from unimportant ones (104, 105). In health communication, a representation that draws attention to important information might thus have larger effects on domain novices, such as new patients learning about a therapy.

3.2 Cognitive load and demands on knowledge

Two related mechanisms underlying some representation effects are:

- *reliance on procedural and domain knowledge*, or the reader's need to know, recall, and apply numeracy skills to use the information. Domain knowledge relevant to numeracy would include recognizing the % symbol or recalling a formula; procedural knowledge includes skills such computing percentages (17).

- *cognitive load requirements*, or the number and difficulty of computations or other mental operations required to use the information productively (17).

In one risk comprehension study, understanding was highest when cognitive load was decreased by completing calculations for the respondents instead of requiring them to calculate (5). In a series of standard gamble studies, higher cognitive effort needed to manipulate the
information about options was associated with more inconsistent decision-making (i.e., more preference reversals) (106). Probability information was presented either in the form of complex-looking fractions such as 513/570, or in simpler-looking but equivalent ones such as 0.90. Participants in the "hard fraction" group rated the task as harder, took longer to respond, and spent a larger proportion of time examining the probability information. However, they spent less time integrating probability and value information, suggesting that this time was not spent transforming the fractions into easier-to-use ones. The researchers proposed that with simple-looking probability information, decision-makers are more likely to calculate or estimate expected values, whereas complex information is more likely to trigger heuristic decision strategies (106).

In the Bayes' Theorem studies described earlier in Chapter 2 (73, 74, 76), the natural frequencies explanation involved adding and comparing integers, and the relevant operations were obvious whether or not the reader knew or recalled Bayes' Theorem. By contrast, the traditional probabilistic explanation required (a) recognizing the problem as a Bayes' Theorem problem, (b) correctly recalling the formula for Bayes' Theorem, and (c) applying the formula by multiplying, dividing, and adding fractions. Mathematical manipulations on fractions are typically more difficult than those on integers (107).

The study in ob-gyn clinics mentioned earlier (22) also involved alterations in cognitive load and domain knowledge. The question "Which rate is higher, 1 in 384 women, or 1 in 112 women?" requires respondents either to perform relatively effortful cognitive manipulations or to draw upon heuristic knowledge. For example, one way to solve the problem would be to calculate that $1 + 384 = 0.0026$ and $1 + 112 = 0.0089$. An easier method would be to reason that a larger denominator means a smaller quantity, but this method requires familiarity with the heuristic of comparing denominators, likely to be acquired through experience with solving math problems.

By contrast, Grimes and Snively's easier version of the same question ("Which rate is higher, 2.6 per 1000 women, or 8.9 per 1000 women?") holds the denominator constant and requires only that the reader compare the two numerators (22). Comparison is simpler than multiplication (107), so the question imposes less of a cognitive burden. (The Grimes and Snively study also probably involves a salience effect: In version 1, the denominators are more salient than the
numerators because they are different from each other. The possibility that respondents with less
education compared the denominators and not the numerators is suggested by the fact that 70%
of the less educated women chose "1 in 384" as smaller than "1 in 112.")

Reducing cognitive effort by using simple calculations or comparisons in place of more
complex operations is likely to be particularly effective for people with low numeracy. Peters has
shown that low-numeracy college students are more strongly affected by framing effects than
high-numeracy ones, probably because the low-numeracy subjects are less likely or able to
mentally translate between representations (55). The low-numeracy subjects were also more
affected by other format effects (for example, making different choices when a risk was
represented as "1 in 10" than as "10").

Graphics can reduce cognitive effort by facilitating visual comparisons between quantities,
rather than cognitively effortful numeric comparisons (108). The ease of visual comparisons is
determined by the number and complexity of mental operations required (109). For example, if a
pair of bar charts had been constructed to show the two risks in the Grimes and Snively study
(22), no mathematical operations would be required to determine which was bigger. Decisions
can be made on the basis of visual readout rather than cognitive processing.

However, the ease and accuracy of visual comparisons depends upon the design of the
graphic. When elements of the graphic are directly proportional to the quantities being compared,
then visual comparison is also linear. A graphic using a log scale (110) has elements that are not
directly proportional to the quantities they represent, and is thus not useful for linear comparisons
of the actual difference between quantities. (However, for technically educated viewers, log
scales are better for calling attention to percentage change or multiplicative differences (79)).
Similarly, truncating the y axis in a bar chart will produce bars that are not proportional to the
quantities being compared (78). Viewers directly comparing the visual elements may reach
misleading conclusions unless they draw upon heuristic knowledge of how to interpret graphs in a
particular domain, or perform mental calculations to reconstruct the quantities being represented
(Figure 3.1). (This figure will also be discussed below in the section on visual perception.)
Figure 3.1: Non-proportional graphics. On the left, the truncated y axis results in bars that are not directly proportional to the quantities being represented (111). The effect is to exaggerate differences between groups (in this case, emphasizing the authors' point that strontium levels were higher in teeth of children living close to nuclear reactors). Extending the y axis to 0 makes the visual elements proportional to the quantities represented (right) and shows that at several of the reactors, the differences were very small.

More broadly, according to the proximity compatibility principle (108), comparisons and computations upon different graphical elements are facilitated by spatial proximity, but also by "perceptual proximity," that is, perceptual similarity. Thus, elements of the same color, or elements with similar orientations, will be easier to group, integrate, or otherwise process.

Different types of animations also seem to have different effects on cognitive load. Animated graphics allow processes to be represented as processes, in what one computer game theorist has dubbed procedural rhetoric (112). For example, the process of using an asthma inhaler can be represented as an animated process, rather than as a series of snapshots presented one after the other. This procedural rhetoric could reduce cognitive effort by showing a process, rather than requiring the viewer to imagine how static descriptions or pictures are linked in a process.

On the other hand, animated graphics sometimes appear to increase cognitive burden. Animated instructions can impede comprehension because they require the viewer to hold a series of past states in memory (95). By contrast, a sequence of drawings of the instruction allows the viewer to consult and compare current and past states as needed. Furthermore, animations can create cognitive problems if they do not match viewers' preexisting mental representations of a process. For example, people tend to think of procedures such as traveling and assembling furniture as series of discrete steps rather than as continuous activities, so instructions broken into discrete steps are more effective than animations or movies (95, 113).
A variety of visual perception issues are known to affect the interpretation of data graphics. Weber's law states that the smallest perceptible change in any stimulus is roughly a constant proportion of the original stimulus (114, 115). Thus, for example, a 1-unit difference between bars of a bar chart will be obvious when the bars are short and 1 unit represents 50% of the longer bar, but the same difference is obscure when the bars are long and 1 unit represents only a small percentage of the bars.

Gestalt principles describe the perception of what is figure and what is background (116, 117). For example, elements clustered in close proximity tend to be perceived as a single figure (proximity), as do elements that resemble each other but are not in proximity (similarity). Continuity breaks along the edges of perceived figures will often not be perceived at all (closure). Gestalt effects make possible configural graphics, graphics designed to facilitate recognition of a pattern that is more important than any of its component data points (Figure 3.2) (79, 80, 118).

Such perceptual issues combine with salience and cognitive load to result in making certain types of scientific illustrations easier or harder to use for various purposes. For example, circles or squares of different sizes and slices of pie charts are difficult to compare accurately (79, 80, 119). By contrast, bars in a bar chart, or dots plotted along a common scale are easier to estimate and compare. Also, differences between curves and angled lines are difficult to estimate accurately, while differences between horizontal lines or between a single horizontal line and an angled line are easier to determine (Figure 3.2). Such graphic features were identified in perceptual studies by Cleveland et al (79, 80). When assembled into a taxonomy by another researcher (119), they predicted performance of relatively educated viewers on a variety of tasks including estimating sizes and comparing quantities from graphs.
3.4 Representations and learning

Animation and interaction provide opportunities for interaction over time, which is likely to have learning effects. Electronic health games often incorporate simple flash-card type drills, as well as virtual rewards and penalties to encourage or discourage specific behaviors (87, 88). One asthma education game allowed the player to steer his or her avatar toward or away from asthma triggers such as cockroaches; approaching the trigger caused the avatar to cough and the screen to dim. Administering rescue meds restored the character and brightened the screen (87, 88). Similar games have focused on self-management of diabetes (87-89) and asthma (90-93).

Animations could also facilitate more complex learning, such as the creation of sophisticated mental representations of a process. A central principle of constructivist learning theory is that learners do not passively receive information but actively construct new knowledge on the basis of both newly encountered information and prior knowledge (120). Learning is enhanced through active participation such as searching for additional information or predicting the results of actions. Although associative learning can promote learned responses to familiar situations, sophisticated mental representations, or mental models, can help people understand complex patterns, predict the results of actions, and cope with unusual situations (105, 121). The development of more sophisticated mental representations of health and disease might be responsible for the success of some interactive health education systems and games. In two web-based diabetes simulators, patients input detailed medical data using form-style input and
received personalized graphs of serum glucose levels and other medical indicators (122, 123). In a small randomized trial, patients who used one of these simulators in addition to participating in a counseling program had better glycemic control and hemoglobin A1c levels than patients in the counseling program alone (122). Emmons and colleagues found that a computer tool that allowed participants to input their own risk factors helped correct misperceptions about colorectal cancer risk (124). An active engagement version that allowed people to see the effects of altering their own modifiable risk factors was linked with more accurate risk perceptions than a passive version, although the effect was not statistically significant (124).

Interactive simulations can affect memory and decision-making by creating powerful virtual experiences that closely replicate real ones. Both spatial knowledge (125) and procedural skills acquired in a virtual environment can translate into the real world. Procedural skill translation has been demonstrated in skills ranging from mental rotation (126-128) to pedestrian crossing skills for children (129) to surgical skills (130, 131).

Animation and interactivity could also affect learning and decision-making by exploiting different mental processes. A variety of dual-process theories posit the existence of qualitatively different mechanisms of decision-making, attitude change, and other psychological processes; the posited mechanisms include a rapid intuitive emotion-driven "hot" level and a more cognitively effortful, "cool," analytic level (42, 66, 132). Visual and experiential information is generally linked with the emotional system. Learning may also be associated with multiple information-processing channels (133-135). Presenting the same information in multiple formats (e.g., verbal/textual and pictorial/graphical) improves learning by involving more than one of these channels (133-135).

One research group has suggested that interactive learning about probability information in a standard gamble study exploits associative learning (32). During the interactions in the card-draw experiments described in the previous section, participants are believed to learn the probability of an event through classical associative learning (32). By contrast, understanding or manipulating probabilities when written as numbers (such as 12 in 100, or 12%) requires using learned and relatively effortful mental operations, which helps explain why deficits in interpreting probabilities are linked with low educational level (20, 81, 82) and low numeracy (18, 19). An intriguing
implication is that interactive presentations of quantitative information might help narrow discrepancies in risk perceptions between high- and low-numeracy readers.

3.5 Chapter summary

Potential mechanisms of representation effects include manipulation of salience, cognitive effort, visual pattern detection, and learning effects. These mechanisms apply to novel representations made possible by computer-mediated communication, such as interactive animations. In light of this literature, and of the combined risk communication model presented earlier in this chapter, further study is warranted on the effect of interactive animations on risk communication. The body of work described here raises the possibility that interactivity and animation could alter patient decision-making in beneficial ways (for example, in improving the understanding of numeric risks, or assisting patients in making affective judgments about quantitative information). It is also possible that it could improve communication by reducing differences between the ways that health care professionals and patients perceive risks.
CHAPTER 4. RESEARCH METHODS

This chapter describes methods for a three-stage research project. First, it describes methods for a systematic literature review on risk graphics in health communication, with results presented as an ontology of risk graphic features and their cognitive effects. Second, this chapter reports methods for a formative qualitative study of the development of novel risk graphics. Third are methods for a questionnaire study to measure the effects of these novel graphics on decision-making and perceived risk.

4.1 Systematic review and ontology

(Ancker, Senathirajah, Kukafka, Starren, JAMIA, 2006)

4.1.1 Objective and research aims

A systematic review seeks to avoid potential bias in the selection or summarization of research evidence by performing comprehensive literature searches, predefining selection and assessment criteria, and publishing details of these procedures as part of the methods. In collaboration with several colleagues (Senathirajah, Kukafka, and Starren), I performed a systematic review of peer-reviewed literature in medicine, psychology, patient education and counseling, allied health, and related fields in order to create a comprehensive survey of types of graphics that are used to communicate health risk to consumers and patients.

To describe the types of risk graphics and the links between these types and behavioral or cognitive outcomes, a systematic classification of the types is needed. Such a systematic classification, or ontology, maps the domain of interest and establishes common naming conventions (136-141). “The goal of philosophical ontology is to provide clear, coherent and rigorously worked out accounts of the basic structures to be found in reality “ (136), p. 7. Chapter 5 presents the results of this systematic review organized as an ontology.

4.1.2 Methods

Literature searching for the systematic review sought to retrieve all research on health risk
29

graphics. For literature published before 1999, I relied upon an excellent review by Lipkus and Hollands (24). In recent years, a number of relevant studies have appeared in the medical, psychological, and patient counseling literature. I therefore updated Lipkus and Hollands' article by performing a systematic review of experimental or focus group research on graphs of quantitative health risks. I searched for evaluation studies of graphs describing probabilities, frequencies, or chances of health events that had not been covered in Lipkus and Hollands' review. I excluded commentaries and instructions (e.g.,(84)), studies of pain scales, utility measures, or illustrations that communicated threat or causal relationships (e.g.,(142)), and studies in which graphics were not used as an independent variable (e.g.,(83, 143, 144)). I also excluded dissertations. I searched 3 bibliographic databases (PsycInfo, MEDLINE, CINAHL) and one portal (ACM Portal) for 1998-2005 inclusive using topic headings chosen from the controlled terminology of each database and additional key words (see Appendix A). I read all titles, then read the abstract and full text of potentially relevant articles. Upon identifying eligible articles, I used the 'find citing research' and 'find similar' tools and searched reference lists. I also did key word searches on the websites of selected journals (Health Psychology, Risk Analysis, and Medical Decision Making). In this review, I describe 24 studies. The searches produced: 969 unique articles on MEDLINE of which 14 met the criteria; 245 on PsycInfo of which 3 met the criteria and had not already been identified; and 54 on the Medical Decision Making web site of which 2 were eligible and had not been identified. The other searches produced no articles that had not been previously identified; 5 articles were identified through bibliography search. Additional details about the searches are presented in Appendix A in this dissertation.

4.2 Qualitative study

4.2.1 Objective and research aims


In this study, I sought to use ideas generated from the literature review to develop and prototype interactive risk communication tools that might be appropriate for Web-based health promotion. The qualitative (formative) assessment was designed to explore consumer preferences for
different interactive graphics, basic usability, and consumer interpretations of the graphics. The findings were also used to make changes in an iterative prototyping process, a type of formative research in software development in which the prototype is updated using user suggestions. Results are presented in Chapter 6.

4.2.2 Methods

I drew from two research traditions: social science qualitative methods with their focus on eliciting interpretations and meaning, and technology development with its focus on creating user-centered products (145). Specifically, I sought to combine the advantages of traditional focus group research with those of a software development method called scenario-based usability engineering (146, 147). In scenario-based usability engineering, focus group participants meet with developers to discuss realistic stories ("scenarios") describing individuals encountering a problem and using a computer system to resolve it. These focus group results are used to develop the specifications for the software system, resolve usability problems, and ensure that the product meets users' needs. A key component of this development method is that problems identified by one group of users are remedied immediately, so that the next group will not encounter the same problems but will instead test the changes and identify new issues. The goal is to maximize the users' contributions, not to conduct a controlled experiment of usability. Small sample sizes are justified by Nielsen's sample size formula for usability testing: 

\[ P = 1 - (1 - p)^n, \]

where \( P \) is proportion of problems found, \( p \) is the probability an individual will find a problem, and \( n \) is the sample size (147). Under the assumption that any individual has a roughly 30% chance of finding a particular usability problem (an assumption based on Nielsen's empirical findings with relatively simple systems), a sample as small as 3 can be expected to identify 67% of usability problems, and a sample of 5 can be expected to catch 84% of problems (147). I thus planned a series of groups of 3-5 people. Keeping the group size small was important for pragmatic reasons: it allowed us to provide a laptop for each participant, ensured that they could get any needed technical help and see each others' computers during discussions, and allowed the moderators to handle the inevitable technical problems.
From a social science perspective, one limitation of scenario-based usability methods is that focus groups are often conducted by software developers themselves, which seems likely to alter the dynamics of the focus group, particularly by inhibiting the participants' willingness to criticize the prototype or concept. I therefore engaged an experienced focus group moderator to conduct the groups. Using a script, she introduced herself as a mediator between the developer (JSA) and the participants, and probed participants' responses to elicit both negative and positive criticism.

The goal was to develop an approach suitable for use with health consumers and patients, who are likely to be heterogeneous in demographics, health status, computer experience, and goals. To recruit members of the public, I partnered with Harlem Health Promotion Center (HHPC), a community-based participatory research center in Harlem, NY, funded by the Centers for Disease Control and Prevention. The goal of community-based participatory research is to bring together community members with academic researchers to ensure that the research serves the community's self-identified needs and goals (148). I worked with HHPC's Digital Partners, a project to examine Harlem residents' use of technology and explore how technology can be used to promote community health goals. I conducted focus groups at HHPC's offices, invited the focus groups to provide feedback on early versions of HHPC's new community health web portal (149), and contributed the risk communication module to the health portal under development.

All discussions were professionally transcribed, and two researchers (JSA and CC) coded the transcripts using a grounded theory approach to content analysis (150). The two coders first reviewed the transcripts and produced tentative lists of codes at the level of the sentence fragment, sentence, or paragraph. We then reached consensus on a codebook. (The objectives of the study were used to develop the focus group facilitator guide, including probe questions asked by the facilitator, but not to develop the concept codes.) The coders then annotated the transcripts individually, and met again to reach consensus on the line-by-line coding of each transcript; codes were added to the codebook if both coders agreed they were needed. Transcripts were annotated using ATLAS.ti™ software (ATLAS.ti, Berlin, DRG). For the consensus discussions, the ATLAS.ti graphical interface was used to clarify the meaning and
relationship of the concept codes by linking them into categories, forming a semantic network or
graphical representation in which nodes represent concepts and links represent relationships
between them (151). For example, codes describing emotional responses were determined to be
negative, positive, or mixed. They were then linked through <is_a> relationships to parent nodes,
e.g., <negative feelings>. A focus group discussion about a topic, such as the quality of doctor-
patient communication, could then be analyzed by examining the intersection between the set of
passages coded with <doctor-patient relationship> and the set coded with <negative feelings>
and its children codes, or <positive feelings> and its children. The sets could also be counted and
several examples are provided in the results section. However, these numbers should be
interpreted cautiously as rough indicators of associations rather than as true quantitative
measurements or survey data (participants were not polled and many expressed multiple
opinions).

Focus groups were invited to discuss a "personalized risk scenario," modeled after stimuli
used in recent studies of graphics (26, 81, 86). In this scenario, a character named "Michelle"
goes to a doctor, who examines her, states her risks of developing heart disease and breast
cancer (e.g., “Your lifetime risk is 46%”), and gives her advice about a heart-healthy lifestyle.
Focus group members were asked to imagine that "Michelle" has asked them for help
understanding the risk. (Two focus groups also discussed a scenario in which a character has the
choice of going unvaccinated and risking a serious disease or taking a vaccine that reduces the
risk of disease but carries a risk of side-effects. However, results are not presented here as most
groups ran out of time before reaching this scenario.)

Participants discussed the scenario and suggested ways to explain the risks to "Michelle." For three of the focus groups, I also created an interactive cardiac risk calculator based on the
National Cholesterol Education Program's guidelines (9) so that participants could input personal
characteristics (age, sex, cholesterol level, etc.) and receive a numeric risk similar to the one in
the Michelle story, as well as explore the effect of changing variables such as cholesterol level.
Participants were also invited to discuss the utility of illustrating the risk with three static graphics:
a simple bar chart, a matrix of stick figures with affected people scattered randomly throughout
the group (Figure 1A), and a matrix of figures with those affected arranged sequentially (Figure 1B).

Participants then used laptops to explore the interactive risk communication module. This module went through several iterations, as described below, but all of them had the same basic structure. In this structure, the first risk presentation screen displayed text and a percentage risk (e.g., “It’s impossible to tell exactly what will happen to you in the future. However, doctors do know what happened to other people just like you. A 6% chance means that in the past, when doctors examined 100 people like you, about 6 of them had a serious heart attack in the next 10 years. So we know how many people will be affected, but not which ones.”). The second screen displayed a graphic (e.g., the random arrangement of stick figures, as in Figure 1A or the sequential arrangement as in Figure 1B). A display with 1000 small figures was selected so that it could show small risks as well as large ones, and so that it would convey the impression that the data came from a large sample of people (26, 81). The colors (pale green and dark blue) were chosen to provide good contrast on the screen and to avoid any potential racial implications of using white, black, or brown. The final screen in the module invited participants to interact with the graphic in some way (as in Figure 2). Participants could then finish or return to the start to explore different risk levels (e.g., by inputting a lower cholesterol level). The user advanced through the program by clicking on large navigation buttons. No text passage was higher than the 9th-grade reading level by the Flesch-Kincaid scale (152) and most were at the 6th-grade level.

The study was approved by the Columbia Institutional Review Board. Participants were recruited through fliers at public clinics, libraries, recreation centers, and other organizations in Harlem, as well as from among contacts from previous studies at the Harlem Health Promotion Center. All provided written informed consent and were reimbursed with movie ticket vouchers and subway passes.
4.3 Questionnaire study

4.3.1 Objective and research aims

This questionnaire study was developed to assess the effects of graphics developed in the formative research project on risk perception and health-related decision-making. Two interactive graphics and two static control graphics were developed on the basis of the qualitative findings (Chapter 6).

**Figure 4.1 Group assignment**

In Group 1, both stories were illustrated with static *random* graphics, that is, blue figures (representing those with disease) and yellow ones (representing those without the disease) were scattered randomly throughout the grid. (www.dbmi.columbia.edu/~jsa7002/dissertation/final/Random.php; requires Adobe Flash Player)

In Group 2, both stories were illustrated with static *sequential* graphics, that is, the blue figures were sequentially arranged in the bottom rows of the grid. (www.dbmi.columbia.edu/~jsa7002/dissertation/final/Sequential.php)

In Group 3, stories were illustrated with interactive *switch* graphics, which allowed participants to toggle back and forth between sequential and random views of the same percentage. Participants were required to switch views at least twice before being able to answer the questions. (Working example at www.dbmi.columbia.edu/~jsa7002/dissertation/final/Switch.php)

In Group 4, stories were illustrated with interactive *search* graphics. In this graphic (described in more detail in section 4.3.2.2 and illustrated in Figure 6.1), participants first saw a grid of orange squares, with instructions to click on any square to see the figure underneath. When a blue figure was found, all the squares turned over to reveal all the figures. Participants had to find a blue figure before they could answer the questions. (Working example at www.dbmi.columbia.edu/~jsa7002/dissertation/final/Search.php)
The primary outcomes were measures of perceived risk and the decision about whether to take protective measures against the threat illustrated with the graphic. Covariates included numeracy, health literacy, and sociodemographics. Secondary outcomes included perceived helpfulness and credibility of the graphics. An additional section of the same study presented unlabeled graphics and asked for estimates of the proportion represented. Results are presented in Chapter 7.

Research Aim A. To explore effects of numeracy on risk perception

Hypothesis A1. Low numeracy will be associated with higher risk perceptions and less accurate estimates with all graphics.

Research Aim B. To compare the effects of interactive and static graphics on perceived risk

Hypothesis B1. Random graphics will be associated with higher perceived risk than sequential ones.

Hypothesis B2. Random graphics will be associated with more variability (lower accuracy) in risk estimates than sequential ones

Hypothesis B3. The interactive switch graphic will produce perceived risk answers midway between the random and sequential ones.

Hypothesis B4. The interactive search graphics will be associated with higher feelings of risk for large risks, and smaller feelings of risk for small risks, than static graphics are.

Research Aim C. To compare the effects of interactive and static graphics on intention to take preventive action

Hypothesis C1. Random graphics will be associated with stronger intention to take preventive action.
Hypothesis C2. The interactive switch graphic will be associated with intention answers midway between those associated with the random and sequential ones.

Hypothesis C3. The interactive search graphics will be associated with stronger intentions to take preventive action for large risks, and weaker ones for small risks, than static graphics.

Research Aim D: To examine interactions between numeracy and effects of the graphics

Hypothesis D1. There will be an interaction between the interactive search graphic and numeracy such that the interactive graphic will be associated with reduced differences in perceived risks between high- and low-numeracy participants.

Hypothesis D2. There will be an interaction between the interactive search graphic and health literacy such that interactive graphics will reduce differences between responses of high- and low-numeracy participants.

4.3.2 Methods

4.3.2.1 Participants

The target population was adult health consumers, so two samples were recruited to ensure a broad range of numeracy, health literacy, and computer literacy levels. One sample was recruited in waiting rooms at NewYork Presbyterian Hospital, an urban tertiary care hospital in a largely Hispanic neighborhood in northern Manhattan. A second on-line sample was recruited by sending emails to eligible volunteer participants in Columbia's Center for the Decision Sciences Virtual Laboratory. The 30,000-member pool contains volunteers from a wide variety of geographic locations, professions, and ages willing to participate in paid questionnaire studies about decision-making. About 30% have a bachelor's degree or higher (personal communication, Ryan Murphy, PhD, associate director, Center for the Decision Sciences). Upon registering with the
participant pool, each member also registers a PayPal account linked to a credit card. This verifies participant identity and ensures that no one can participate more than once in a particular study.

The questionnaire study was approved by the Columbia Institutional Review Board (IRB-AAAB8444). All participants granted informed consent in an on-line form and were provided with a small monetary incentive upon completion (movie ticket voucher for clinic participants, equivalent dollar amount through PayPal for lab participants). Data were stored in a password-protected database. The PayPal account information was stored separately from all responses.

4.3.2.2 Graphics

All graphics were developed in Adobe Flash CS Professional v 9.0, using ActionScript 2.0 (Adobe Systems Inc., San Jose, CA) and embedded in an html/php questionnaire that passed data to a MySQL database. All graphics used a grid of stick figures with yellow indicating no disease and dark blue indicating disease. A 24 x 12 size was chosen to carry the implication of a large sample; small stick figure samples are sometimes interpreted as indicating small and therefore unreliable sample sizes (81). Also, I avoided 100 in order to discourage counting and encourage a gestalt impression of the graphics, particularly for the estimating substudy of unlabeled graphics (section 4.3.2.4 below).

The design process raised questions about the best interpretation of random in this context. For the random graphic, a binomial distribution could be produced by attaching a random number generator to each stick figure to ensure it would turn blue with a probability of exactly 29% for Story 1 and exactly 6% for Story 2. However, this design has two drawbacks. First, each participant would see a different random arrangement, which could cause unwanted variability on the perceptual level. Also, each participant would see a different proportion of blue figures. For example, for the 6% graphic, the expected value under the binomial distribution would be \( Np = 240 \times 0.06 \approx 14 \) stick figures, but the variance would be \( Np(1 - p) = 240 \times 0.06 \times 0.94 = 14 \). Thus, roughly 16% of participants might be expected to see 10 or fewer stick figures (a proportion of \( 1/240 = 4\% \)) and another 16% might see 17 or more figures (a proportion of \( 17/240 = 7\% \)).
Instead, I set the proportion of blue stick figures at exactly 29% for Story 1 and 6% for Story 2, and wrote a permutation algorithm that created a single random arrangement for each story. The resulting illustration was used for all participants in the *random* graphic group.

The same illustration was used in the *search* graphic. In consequence, the *search* interaction used sampling without replacement (hypergeometric distribution) rather than sampling with replacement (binomial distribution). Thus, the chance of hitting a blue person was not always exactly 6% or 29% but rather changed slightly from click to click. To illustrate, let $N$ be the size of entire collection of stick figures, containing $m$ blue figures and $N-m$ yellow figures. Let $n$ equal the sample size (i.e., number of squares clicked by the viewer) and $k$ equal the number of blue stick figures in the sample. The hypergeometric probability distribution is:

**Equation 4.1:** \[ pdf_{\text{hyper}}(k, N, m, n) = \binom{m}{k} \binom{N-m}{n-k} \binom{N}{n} \]

Furthermore, the 'game' ends when a blue person is discovered. The probability that $k_n = 1$ (i.e., of reaching $n$ clicks before hitting a blue person) is described with Equation 4.2.

**Equation 4.2:** \[ p(k_n = 1) = pdf_{\text{hyper}}(1, \ N - [n - 1], \ m, \ 1 \ | \ 0, \ N, \ m, \ n-1) \]

This equation breaks the probability into two independent events: the chance that a sample of size 1 will hit 1 blue person in a reduced collection of $N - n + 1$ stick figures including $m$ blue ones; and the chance that a sample of size $n - 1$ will hit no blue people in the entire collection of $N$ stick figures and $m$ blue figures. Because these events are independent, this is equivalent to:

**Equation 4.3:** \[ p(k_n = 1) = pdf_{\text{hyper}}(1, \ N - [n - 1], \ m, \ 1) \times pdf_{\text{hyper}}(0, \ N, \ m, \ n-1) \]

The first half of the equation represents the probability of hitting a blue person on one click if all previous clicks have hit yellow people. Table 4.1 shows the probability distribution for the 6% graphic, demonstrating that the per-click probabilities vary very little from 6%.
Table 4.1: Hypergeometric probabilities of numbers of clicks in the search graphic

<table>
<thead>
<tr>
<th>Number of clicks (n)</th>
<th>Chance of hitting blue after n-1 yellow clicks*</th>
<th>Chance of reaching exactly n^a</th>
<th>Cumulative probability of hitting blue by n^b click*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0583^d</td>
<td>0.0583</td>
<td>0.0583</td>
</tr>
<tr>
<td>2</td>
<td>0.0585</td>
<td>0.0551</td>
<td>0.1134</td>
</tr>
<tr>
<td>3</td>
<td>0.0588</td>
<td>0.0521</td>
<td>0.1656</td>
</tr>
<tr>
<td>4</td>
<td>0.0590</td>
<td>0.0492</td>
<td>0.2148</td>
</tr>
<tr>
<td>5</td>
<td>0.0593</td>
<td>0.0465</td>
<td>0.2614</td>
</tr>
<tr>
<td>6</td>
<td>0.0595</td>
<td>0.0439</td>
<td>0.3054</td>
</tr>
<tr>
<td>7</td>
<td>0.0598</td>
<td>0.0415</td>
<td>0.3470</td>
</tr>
<tr>
<td>8</td>
<td>0.0600</td>
<td>0.0392</td>
<td>0.3862</td>
</tr>
<tr>
<td>9</td>
<td>0.0603</td>
<td>0.0370</td>
<td>0.4232</td>
</tr>
<tr>
<td>10</td>
<td>0.0606</td>
<td>0.0349</td>
<td>0.4582</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>20</td>
<td>0.0633</td>
<td>0.0192</td>
<td>0.7146</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>30</td>
<td>0.0663</td>
<td>0.0103</td>
<td>0.8542</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>50</td>
<td>0.0732</td>
<td>0.0027</td>
<td>0.9657</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>226 (i.e., 240-14)</td>
<td>100.0</td>
<td>&lt;0.0001</td>
<td>100.0</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>240</td>
<td>0.0583</td>
<td>0.0000</td>
<td>100.00</td>
</tr>
</tbody>
</table>

a. \( p = \text{pdf}_{\text{hyper}}(1, N - [n - 1], m, 1) \)
b. \( p = \text{pdf}_{\text{hyper}}(1, N - [n - 1], m, 1) \times \text{pdf}_{\text{hyper}}(0, N, m, n-1) \)
c. \( p = \sum \text{pdf}_{\text{hyper}}(1, N - [n - 1], m, 1) \times \text{pdf}_{\text{hyper}}(0, N, m, n-1) \)
d. Graphic displayed 14/240 = 0.0583 rather than 15/240 = 0.0625.

4.3.2.3 Scenarios and instrument

All participants read two short scenarios in which they were asked to imagine that their doctor, "Dr. Smith," had said that they had a particular risk of a disease, and had offered the option of a free protective medication or vaccine (Appendix B). In Story 1, the probability of a flu-like disease was 29%, and in Story 2, the probability of developing heart disease was 6%. The preventive medication was very effective, but in order to avoid ceiling effects, I described it as having some side effects so that not everybody would choose it. The chance of side effects was about 1/3 the probability of disease without the preventive action (i.e., 9% for the 29% story, and 2% for the 6% story).

Each story was followed by 11 questions designed to assess perceived risk, intention to take preventive action, constructs from EPPM (self-efficacy, response efficacy of the protective measure, perceived severity of the disease) (3) and HBM (barriers/drawbacks to the protective measure) (1), and realism (26). The realism question was suggested by the qualitative research findings (Chapter 6), in which random graphics were repeatedly described as more "realistic" than
sequential ones, as well as by Schapira's exploration of the perceived "truth" of different graphics (26). These concepts of realism and truth may be associated with credibility of the information.

As risk perception was of primary interest in this study, I used 4 questions to assess it. Two questions about perceived susceptibility and perceived vulnerability were adopted from Weinstein (65), who validated these questions against subsequent health-protective behavior (flu vaccination). A third question (verbal risk estimate) asked for a risk description on a 7-point verbal scale from "almost zero" to "almost certain", and a numeric estimate of their own risk on a scale from 0 to 100% (numeric risk estimate), adapted from questions used to assess breast cancer risk perceptions (153). The other 11 items used a four-point forced-choice response scale used by (65): agree strongly, agree mildly, disagree mildly, disagree strongly.

Health literacy was assessed with the Short Test of Functional Health Literacy in Adults (S-TOFHLA), a validated test of health literacy that requires approximately 7 minutes to complete (70, 71). The S-TOFHLA assesses comprehension of excerpts from materials such as patient education brochures and prescription labels, unlike shorter reading tests such as the REALM (154), which simply require pronouncing medical words aloud (15). This instrument has 36 cloze procedure exercises, but participants answer only as many as they can complete in 7 minutes. The score is used to assign respondents to a literacy category ('adequate,' 'marginal,' or 'low'). I designed an on-line version of the S-TOFHLA that resembled the print version as closely as possible, but it has not been validated in the on-line format.

Numeracy was assessed with an 8-item scale developed by Lipkus, which has been normed in an on-line population (19).

Sociodemographics were collected on the final pages of the survey. Computer use was assessed with three questions: comfort using the mouse from (155), frequency of use, and presence/absence of email address. (The mouse question was not used in the analysis as most of those who reported difficulty using the mouse were daily computer users, perhaps indicating repetitive stress problems rather than lack of familiarity with computer technology.)

Validity checks were added to force completion of the risk perception, decision, and demographics questions. Literacy and numeracy questions could be left blank (blank responses
were counted as wrong). Half of respondents in each group got questions in reverse order to test question order effects.

### 4.3.2.4 Substudy on unlabeled graphics

All participants were also shown a sequence of 6 unlabeled graphics and asked to estimate the proportion of blue stick figures in each on a scale from 0 to 100%. Each graphic faded to gray after 10 seconds to ensure a uniform deadline and to discourage counting. All participants saw illustrations of 6% and 29% in both sequential and random arrangements, as well as another graphic in sequential and random arrangements (either 40%, 50%, 60%, or 70%). The unlabeled graphics section was placed before the scenarios study.

### 4.3.2.5 Pilot testing and reliability

The on-line questionnaire was piloted with 7 participants recruited by word of mouth from among NewYork Presbyterian Hospital support and housekeeping staff to represent a range of education levels (4 high school graduate or less; 3 with college degrees or higher degrees) and computer familiarity (2 described themselves as having little experience with computers, and neither had an email address). The pilot testing was designed to assess usability and comprehensibility, estimate the duration and perceived burden of the questionnaire, and check the technical functionality. It led to clarification of some user instructions and a redesign to reduce the number of questions to 2 per page for most pages, except for the S-TOFHLA pages (up to 5 per page), numeracy (8 items on 1 page), and sociodemographics (6-7 per page).

Observation of the low-computer-literacy pilot testers suggested that they needed longer to complete questions simply because of unfamiliarity with the mouse. As a result, I designed the S-TOFHLA section to record the 7-minute deadline but not to time out until 10 minutes. However, the final data set showed that no respondent was assigned to a different literacy category as a result of the extra 3 minutes, so the original 7-minute score and category is used for all analyses.

Test-retest reliability was assessed by having an additional 9 people (not involved in the focus groups or the pilot) take the questionnaire twice separated by an interval of about 2 weeks.
4.3.2.6 Data analysis

The primary outcome measures were the risk perception questions and intention (or decision). The total sample size of 165 (roughly 41 in each group) provided 80% power to detect between-group differences of 0.75 of the within-group standard deviation at the 0.05 level. Numeracy, health literacy, and sociodemographics (including clinic status vs. online sample) were the most important covariates. Secondary outcome measures of interest were participants' attitudes toward the graphics (perceived as helpful, perceived as confusing, perceived accuracy and realism). Other predictors of intention (mediators) were constructs from EPPM and the health belief model (severity, barriers, response efficacy, self-efficacy). Chi-squared tests and ANOVA were performed across person, to compare responses of different people reading the same scenario but viewing different graphics. Repeated measures ANOVA was performed to assess within-person repeated measures (e.g., responses on the 29% story and on the 6% story). After univariate exploration of the data, general linear models were constructed to predict risk perception and intention. Skewed data were normalized by square-root transformations or tested with non-parametric tests, but results did not differ substantially, so most analyses are presented with non-transformed data and parametric tests. Tests were two-tailed. Analyses were performed in SAS 9.1 and SPSS 16.0.

4.3.2.7 Cognitive analysis

Two additional respondents completed concurrent think-aloud protocols (continuously verbalizing their thoughts) while completing the questionnaire (156). Interactions and audio were captured with Morae software (TechSmith, Okemos, MI). The recordings were coded with a simplified coding scheme that included verbal references and mouse gestures to text, numbers, and graphs (intended to capture salience of different sources of information); and inferences or elaborations from the above.
CHAPTER 5. RISK COMMUNICATION ONTOLOGY

Graphics section based on systematic literature review previously published as (23). Number format section based upon (17).

5.1 Introduction and scope

Risk communications can incorporate words, numbers, illustrations, or some combination of these modalities.

Table 5.1: Examples of risk descriptions using words and numbers

<table>
<thead>
<tr>
<th>Example Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;You can get hepatitis from sharing needles.&quot;</td>
</tr>
<tr>
<td>&quot;Your chance of heart disease is higher than average because you have diabetes.&quot;</td>
</tr>
<tr>
<td>&quot;BRCA1/2 mutations increase the lifetime risk of breast cancer by a factor of 4.&quot;</td>
</tr>
<tr>
<td>&quot;BRCA1/2 mutations are associated with an 80% lifetime risk of breast cancer.&quot;</td>
</tr>
<tr>
<td>&quot;People who take aspirin daily have a 3% chance of developing a stomach bleed: of those with bleeds, 10% die.&quot;</td>
</tr>
<tr>
<td>&quot;The medication increases the chance of side effects (RR: 2.9; 95% CI: 1.7-6.2; p=.02).&quot;</td>
</tr>
</tbody>
</table>

Verbal descriptions of risks have long been known to affect risk perception, but my systematic review of literature on risk graphics (23) also confirms that design features of graphics can affect risk perception, comprehension, cognition, attitudes, and behavior. My review article on numeracy with David Kaufman subsequently identified strong cognitive effects of different number formats, such as whether a risk is depicted as 33%, 33 in 100, or 1/3 (17). This phenomenon, in which presenting the same information in different formats affects comprehension, problem-solving, or decision-making, is called a representation effect (17, 68). Mapping representation effects to features of risk presentations would support predictions about the effects of different risk presentations and help identify hypotheses for future research. Such a mapping would be particularly useful for communicators seeking guidance for ways to express quantitative risks persuasively (as in health promotion) or non-persuasively (e.g., showing evidence for informed decision-making, public discourse, or informed consent).
To ensure that the mapping is both consistent and exhaustive, I first need a description of the types of risk communications and the relationships between them. Such a systematic classification, or ontology, "provide[s] clear, coherent and rigorously worked out accounts of the basic structures to be found in reality " (136). It creates a logically consistent map of the concepts in the domain of interest, establishes common naming conventions, and makes domain assumptions explicit so that information about the domain can be shared and reused (136-141).

This chapter presents an ontology to define and classify risk communication messages, with a particular focus on risk graphics and risk numbers. Where possible, it links the types to known representation effects. The ontology categories are based on evidence about cognitive, perceptual, or decision-making effects of the format of numbers or graphics. For example, ample evidence suggests that comprehension is affected by whether a risk is depicted as a frequency or a percentage (17), but I have found no studies to suggest that numbers as words (e.g., "four") have any different effects from numbers as digits. Thus, the ontology (Section 5.4 below) distinguishes between frequencies and percentages, but not between digits and number words.
Figure 5.2: Ontology of risk communication messages. Solid arrows represent required (essential) components and properties. Dashed arrows represent optional (non-essential) components and properties. Concept map developed in CMap Tools (Institute for Human and Machine Cognition, Pensacola, FL). Some terminal leaves have been pruned for clarity; they are described in the text.
5.2 Four purposes of risk communication

Risk communications in general, and graphics in particular, have been designed for a variety of different purposes (119, 157-161). As I shall discuss in section 5.5, different graphics have been studied in light of different goals, and various graphical features appear to support different goals. I propose four categories of goals: information, education, neutral decision support, and behavior change or persuasion. These goals are not mutually exclusive. For example, ability to convey information is often measured as part of both persuasion and decision support studies.

Figure 5.2 (Detail)

1. Information or knowledge

Many communication campaigns and risk graphics are assessed by asking subjects to extract information from the message. Informational tasks relevant to graphics include:

a. local comparison (e.g., which risk is higher?)

b. point reading (e.g., what is the risk?)

c. proportion estimate (e.g., what proportion of x is y?)

d. visual summation (e.g., is A+B bigger than C+D?)

d. cognitive comparison (e.g., how much higher is this risk than that risk?)

e. synthesis (e.g., what is the trend in risk over time? which option cuts my risk the most?)

Outcome measures used in this type of research include accuracy or consistency of quantitative reasoning or perceptions (e.g.,(30, 78, 80, 162)). Examples from this review include whether users can estimate a proportion represented in a risk graphic (162) and whether they can read the
number of survivors at a particular time point from a survival curve (61).

Carswell (119) and (159) proposed four of these information tasks: local comparisons, point reading, global comparisons (e.g., is A+B > C+D), and synthesis. However, later work with pie charts and bar charts (163) supports making a distinction between proportion estimates that use automated visual perception (e.g., what proportion of the pie is represented by this pie slice?) and those that require visual summation of spatially separated visual elements such as individual bars in a bar chart (a relatively effortful and less accurate mental process). Furthermore, some comparison tasks require both perception and numeracy (e.g., how much higher is A than B?).

2. Education

Many graphics have been assessed in light of whether they promote more in-depth learning of information. Educational interventions may measure the tasks described above as 'information/knowledge" tasks, as well as ability to make more complicated inferences, to predict events or phenomena not explicitly described in the educational materials, or to retain or recall information over time (133, 164).

3. Neutral decision support

Neutral decision support communications are assessed by whether they increase a decision-maker's knowledge about choices and ability to compare them. Examples include descriptions of risks and benefits for cancer treatment options or for informed consent for potential research participants. Neutral decision support interventions may be assessed by assessing information (above) and knowledge, or through factors such as decisional conflict or the match between the option and the decision-maker's previously expressed priorities. For example, one study examined whether viewers produced the same ranking of risks when expressing their perceptions in different graphical formats (165, 166). From this perspective, graphical elements that cause perceptions of the risk to deviate from the probability of the outcome (including framing, axis distortion, or relative comparisons) may be considered unethical (30, 78, 167-170).
4. Persuasion or behavior change

Persuasive or behavior change communications are intended to promote actions, intentions, or perceptions specified by the risk communicator. For example, a message or graphic may be intended to improve medication guideline adherence (actions), promote intention to adopt a healthy behavior (intentions), or reduce framing effects (perceptions). Some persuasive risk communications are designed to increase concern (e.g., raising awareness of important health hazards), while others are designed to calm fears (e.g., reducing unreasonable anxiety about rare risks) (171). Health promotion specialists seeking to induce behavior change may exploit framing and salience effects (44) and consider them justified by their effectiveness on desirable outcomes (172). An example involving graphics is a series of studies of users' willingness to pay (a behavior) for hypothetical consumer products after viewing graphic displays of safety risks associated with each (27).

Some risk communication studies ask participants to rate the perceived attractiveness, helpfulness, or clarity of the graphics. Measuring users' likes and dislikes in this way is believed to be important because in real-world settings, people may not accept, attend to, or remember graphics or communications they dislike (81, 86, 173). This aspect of the graphic may be particularly important to persuasive goals, as deeper processing of the message is associated with more stable and persistent attitude change (66). Although perceived attractiveness is frequently measured, I have not labeled it as a goal in itself as it is generally considered a mediator or proxy for another goal such as education or persuasion.

5.3 Risk descriptions

A risk communication message for any of these four purposes contains a risk description.
1. A risk description has two essential (139) components: a hazard and a condition. In other words, every risk description describes the risk of some hazard happening to a particular group of people or under particular circumstances.

(The term hazard encompasses both of the two EPPM dimensions of "severity" and "susceptibility," and the term condition describes the reader or audience targeted by the message. The advantage of this terminology is that it can be used to describe informational, educational, or decision-support communication as well; for example, condition would describe a particular choice option, such as Treatment A or Treatment B, and hazard would describe consequences of the option such as side effects.)

2. The condition description may be highly specific or very general. "You can get hepatitis from sharing needles" states a specific condition of needle sharing. "There's a 15% chance of getting flu this year" describes only the relatively general condition of "this year."

3. The hazard description has two non-essential components
A. Severity may or may not be described. (That is, the statement may describe consequences of heart disease, or it may simply describe a chance of heart disease, leaving severity to be inferred.)

B. Probability may or may not be described. Hazard descriptions that have no probability are non-quantitative (e.g., "You can get hepatitis from sharing needles").

4. Probability descriptions have two essential properties

A. Level of measurement (left-hand column of Table 4.2). Level of measurement is either semi-quantitative (e.g., "high") or quantitative (e.g., 5%).

B. Relativity: Probability descriptions may be either relative or absolute

5. Probability descriptions have one non-essential property.

A. Time. Probability may have no time property (e.g., "Your risk of stroke is twice as high if you are a smoker"). Alternately, probability may be described as a trend over time (e.g., "Your risk will increase slowly as you age"), or cumulative over a time period ("Lifetime risk of cancer is 40").

Figure 5.3: Coding for sample risk descriptions

<table>
<thead>
<tr>
<th>EXAMPLE 1: &quot;Your chance of heart disease is higher than average because you have diabetes.&quot;</th>
<th>EXAMPLE 2: &quot;BRCA1/2 mutations increase the lifetime risk of breast cancer by a factor of 4.&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;Condition: has diabetes&gt;</td>
<td>&lt;Condition: has BRCA1/2 mutation&gt;</td>
</tr>
<tr>
<td>&lt;Hazard: heart disease&gt;</td>
<td>&lt;Hazard: breast cancer&gt;</td>
</tr>
<tr>
<td>&lt;Severity is not explicitly described&gt;</td>
<td>&lt;Severity is not explicitly described&gt;</td>
</tr>
<tr>
<td>&lt;Probability: higher than average&gt;</td>
<td>&lt;Probability: factor of 4&gt;</td>
</tr>
<tr>
<td>&lt;Level of measurement: semi-quantitative&gt;</td>
<td>&lt;Level of measurement: quantitative&gt;</td>
</tr>
<tr>
<td>&lt;Relativity: relative&gt;</td>
<td>&lt;Relativity: relative&gt;</td>
</tr>
<tr>
<td>&lt;Time is absent&gt;</td>
<td>&lt;Time: cumulative (&quot;lifetime&quot;)&gt;</td>
</tr>
</tbody>
</table>

5.3.1 Probability as component of risk description
As discussed, hazards in risk descriptions may or may not be described with a probability. If probability is used, it has two essential properties: level of measurement (either semi-quantitative or quantitative); and relativity (either absolute or relative to a comparison group). Probability also has one non-essential property: time (either cumulative or trend). The various combinations of these properties lead to a classification that can be portrayed as a tree (Figure 5.3) to avoid the potential problems of multiple inheritance (136) supported by a tabular presentation.

Figure 5.4: Types of probability descriptions in health risk communication

<table>
<thead>
<tr>
<th>Probability</th>
<th>-quantitative (&quot;The chance is 5%&quot;)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>--absolute (&quot;The chance is 5%&quot;)</td>
</tr>
<tr>
<td></td>
<td>--trend over time (e.g., a survival curve)</td>
</tr>
<tr>
<td></td>
<td>--cumulative (e.g., &quot;a 1 in 9 lifetime chance of breast cancer&quot;)</td>
</tr>
<tr>
<td></td>
<td>--relative (e.g., &quot;diabetes doubles your chance of heart disease&quot;)</td>
</tr>
<tr>
<td></td>
<td>--trend over time (e.g., &quot;your risk is increasing as you age&quot;)</td>
</tr>
<tr>
<td></td>
<td>--cumulative (e.g., &quot;smokers have twice the lifetime risk&quot;)</td>
</tr>
<tr>
<td></td>
<td>-semiquantitative (e.g., &quot;The chance is high&quot;)</td>
</tr>
<tr>
<td></td>
<td>--absolute (&quot;The chance is high&quot;)</td>
</tr>
<tr>
<td></td>
<td>--trend over time (e.g., &quot;The chance is increasing&quot;)</td>
</tr>
<tr>
<td></td>
<td>--cumulative (&quot;Your 10-year risk of breast cancer is very low&quot;)</td>
</tr>
<tr>
<td></td>
<td>--relative (&quot;Your risk is higher than average&quot;)</td>
</tr>
<tr>
<td></td>
<td>--trend over time (&quot;Your risk will become higher than average&quot;)</td>
</tr>
<tr>
<td></td>
<td>--cumulative (&quot;Your chance of getting flu this season is relatively high&quot;)</td>
</tr>
</tbody>
</table>

5.3.2 Cognitive effects of probability formats

5.3.2.1 Level of measurement

"Sharing needles is dangerous" or "AIDS can affect anybody" are examples of non-quantitative risk descriptions. If the probability of a health event is mentioned in a risk description, it may be described either semi-quantitatively ("There’s a small chance this is cancer") or quantitatively ("There’s an 8% chance of side-effects with this drug").

Health numeracy studies suggest that many patients have difficulty interpreting or manipulating percentages and other quantitative risk descriptions (17-19, 22, 53-56). In the validation of a health numeracy assessment on a highly educated on-line sample, only 32% got all questions correct, and 16%-20% were unable to answer simple questions such as "Which represents the larger risk: 1%, 5%, or 10%?" (19).

However, semi-quantitative risk descriptions such as "high" and "low" also create problems because their vagueness allows them to be interpreted very differently in different situations (174,
Furthermore, high-numeracy consumers prefer and trust quantitative medical information more than non-quantitative information; conversely, however, low-numeracy consumers placed more trust in communications containing only verbal descriptions (20).

### 5.3.2.2 Relativity

Probability descriptions may be given in absolute terms (quantitative: "a 10% chance"; semi-quantitative: "a high chance") or relative terms (quantitative: "smokers are twice as likely to die from cancer as non-smokers"; semi-quantitative: "a higher-than-average chance").

When people estimate their own personal risk of an illness, they tend to have a difficult time providing an absolute risk, and their relative or comparative estimates are better correlated with objective risk factors than their absolute estimates (21). Providing patients with absolute risks produced a better improvement in self-estimated absolute risk, whereas giving patients a combination of absolute and relative risks produced a better improvement in self-estimated relative risk (124). Worry about cancer is related independently to both absolute risk perception (the individual's estimated likelihood of getting cancer, measured on a qualitative scale from "very low" to "very high") and relative risk perception (how the individual believes his or her risk compares to that of others) (176).

A large body of research confirms that showing relative differences without absolute risks for context inflates the apparent magnitude of the effect, for both expert and novice audiences (168, 169, 177). That is, an increase from 1% to 2% seems larger when it is described as a 100% (relative) increase than when it is described as a 1% (absolute) increase. Adding baseline risk to either absolute risk reductions or relative ones strongly improves the accuracy of interpretation of risk reduction information (178). Understanding of risk reduction information is also improved by completing the calculation for the reader, so that the relative change (such as "this medicine will halve your risk") is accompanied by the resulting risk level (e.g., "this medicine will halve your risk, from 10% to 5%") (5). Overall, this research suggests that knowledge outcomes are best when readers are given all three values: baseline risk, the relative change, and the resulting risk.
5.3.2.3 Time

Risks may or may not be described in terms of time. If time is mentioned, the risk may be either cumulative over time ("your lifetime risk is...") or a trend over time (as in survival curve data, or median survival from cancer, or annual risk of flu). In general, hazards or benefits occurring at a distant time tend to be discounted relative to events occurring immediately (179, 180). Readers' interpretations of risk trends over time have been studied in several experiments (12, 25, 61, 62, 181). In a classic study, McNeil et al provided numeric information about a surgical treatment with high short-term mortality but good median and long-term survival and an alternative radiation treatment with good short-term but worse median and long-term survival (12). When information was framed in terms of mortality, more people chose radiation, perhaps because the short-term mortality appeared more salient in this frame. Framing effects were nearly as strong with physicians as with laypeople. (Survival and mortality curves will be discussed again in the section on graphs.)

5.4 Types of probability numbers

When the hazard component of a risk description has a probability, and this probability is quantitative (whether absolute or relative), the probability may be expressed in any one of several formats of numbers or their verbal equivalents (Figure 5.4): fractions (e.g., 1/4), decimals (0.25), percentages (25%) and frequencies. Frequencies can be subdivided into simple counts (number of deaths), ratios (1 in 4, or 1 of 4 patients will die), odds (1 death to 3 survivors), or rates (250 per thousand). The natural frequencies category is a subset of the ratio category that can be used when the communication goal is to compare several probabilities or perform computations upon them. In a natural frequencies presentation, all the probabilities under discussion are projected onto a single large number, producing a series of ratios with the same denominator. For example, in a ratio presentation, two risks would be described as 1/4 and 1/3, whereas in a natural frequencies presentation, they would be described as 25 in 100 and 33 in 100.
5.4.1 Cognitive effects of probability number formats

Different formats for numbers and quantities strongly affect comprehension, problem-solving, and decision-making (17). As discussed in Chapter 3, these representation effects appear to be mediated by:
salience, or the representation's ability to draw attention to certain quantities or relationships, and minimize attention to others (27-29);

reliance on procedural and domain knowledge, or the reader's need to know, recall, and apply numeracy skills to use the information. Domain knowledge relevant to numeracy might include recognizing the % symbol or recalling Bayes' Theorem; procedural knowledge includes skills such as computing percentages (17).

cognitive load requirements, or the number and difficulty of computations or other mental operations required to use the information productively (17, 182). Cognitive effort has been proposed to consist of a series of elementary processes such as reading, comparing, adding, etc. (182). In math, certain operations such as comparison are simpler than others such as multiplication (107).

In numerical reasoning research, a debate continues about whether frequencies or percentages/decimals are easier to process. Conflicting results in the literature may be due to presentations that differentially triggered salience, domain knowledge, and cognitive load effects.

The "1-in-n" format for frequencies, in which different ratios have different denominators, is universally found to be difficult to use. In a study of patients in a clinic waiting room (22), 73% correctly indicated that a disease affecting "2.6 per 1000 women" was less common than one affecting "8.9 per 1000 women" (a comparison of two rates). However, when the same comparison was presented as ratios with different denominators, i.e., "1 in 384 women" and "1 in 112 women," only 56% answered correctly, suggesting that the rate presentation was easier to understand. The effect was even stronger among patients with low educational level (< 9th grade education), suggesting an interaction between number format and cognitive skills. This study clearly involved effects on cognitive load and domain knowledge. The question "Which rate is higher, 1 in 384 women, or 1 in 112 women?" requires respondents either to perform relatively effortful cognitive manipulations or to draw upon heuristic knowledge. For example, one way to solve the problem would be to calculate that $1 \div 384 = 0.0026$ and $1 \div 112 = 0.0089$. A heuristic method would be to reason that with ratios and fractions, a larger denominator means a smaller quantity. However, this method requires familiarity with the heuristic of comparing denominators,
which is probably acquired through experience with solving math problems. By contrast, Grimes and Snively's easier version of the same question ("Which rate is higher, 2.6 per 1000 women, or 8.9 per 1000 women?") holds the denominator constant and requires only that the reader compare the two numerators (22). This study also probably involves a salience effect: In version 1, the denominators are more salient than the numerators because they are different from each other. The possibility that some respondents, particularly those with low educational level, compared the denominators and not the numerators is suggested by the fact that 70% of the low-education women selected "1 in 384" as smaller than "1 in 112."

The "1-in-n" format also led to the lowest rate of accurate answers in a series of studies with an on-line population (183), while percentages and frequencies of the "1 in 100" format led to higher and similar accuracy rates. Interestingly, when required to add two probabilities, people were more accurate with percentages, but when required to multiply (such as halving or doubling) probabilities, people were more accurate with frequencies (183).

In one study, percentages such as 46% led to more accurate answers than ratios such as 46 in 100 (5). People who received percentage risk information, including how a drug would lower or raise their risk, answered comprehension questions more accurately than people who received frequency information. The difference in accuracy rates was modest (67.4% and 63.4%, p < .01). These researchers and others (184) have argued that the salience of the denominator can interfere with accuracy in the frequency format (5).

Formats in which the information designer completes the calculations for the reader reduces the cognitive effort required to process probability information and improves accuracy in making risk tradeoffs (5). For example, instead of merely stating that a drug will triple the risk of some bad outcome, the communication can present the baseline risk of 4% and the end risk of 12%, and instead of expecting consumers to add the risks of two separate outcomes to decide about the overall usefulness of the drug, the communication can provide the total summed risk for both the outcomes (5).

When the required cognitive load is too high, decision-makers may skip time-intensive processing altogether and use heuristics instead. Complex-looking ratios such as 513/570 are
more demanding to process than equivalent but simpler ones (such as 9/10) or decimals (e.g., 0.90), according to evidence such as self-report and preference reversals with the different formats (106). In a series of standard gamble studies, probability information was presented either in the form of complex-looking fractions or in simpler-looking equivalent ones. Participants in the “hard fraction” group rated the task as harder, took longer to respond, and spent a larger proportion of time examining the probability information. However, they spent less time integrating probability and value information, suggesting that this time was not spent transforming the fractions into easier-to-use ones. The researchers proposed that with simple-looking probability information, decision-makers are more likely to calculate or estimate to produce expected values, whereas complex information is more likely to trigger heuristic decision strategies (106).

In general, the limitations of the “1-in-n” format are not found in the “natural frequencies” presentation, in which denominators of all the ratios in the message are the same large number such as 100, 1000 or 10,000 (73, 74, 76). In one set of experiments, medical students and medical school faculty were given information about the sensitivity and specificity of a diagnostic test, and the prevalence of disease, to compute the positive predictive value of a diagnostic test finding in a particular patient (73) (see Figure 2.3 in Chapter 2). Of the 24 participants who received the information in percentage format, only 1 computed the correct answer; by contrast, of the 24 who received the information as natural frequencies, 16 gave the correct answer (73). The traditional probabilistic format required (a) recognizing the problem as one involving Bayes’ Theorem, (b) recalling the formula for Bayes’ Theorem, and (c) applying the formula by multiplying, dividing, and adding fractions. Mathematic manipulations on fractions are typically more difficult than those on integers (107). By contrast, the natural frequencies explanation involved a lighter cognitive load because it primarily involved adding and comparing integers; there was less demand on domain knowledge because the relevant operations were obvious whether or not the reader knew or recalled Bayes’ Theorem. An evolutionary psychology argument has been made that natural frequencies are easier to manipulate because they require only counting and basic arithmetic, which are considered closer to primitive human reasoning capabilities, and do not require using the percentage format, which is a learned convention (77).
Representational fluency is the ability to translate between or recognize the equivalence of different presentations of the same information (23, 68). Subjects with poor skills in probability reasoning are likely to make different decisions when numbers were presented in different formats (10 in 100 versus 10%), apparently because they are less able to translate between representations of the same quantity (55). The framing effect is probably an example of a representation effect as well. In the framing effect (as discussed in Chapter 2 above), decision-makers make different choices when options are presented in the gain frame (e.g., 90% success rate) than in the loss frame (e.g., 10% failure rate). Framing effects are nearly universal and have been detected in physicians as well as patients (12). However, poor numeracy skills are associated with greater susceptibility to framing effects, possibly because of the association with representation effects (55). There is some evidence that presenting information in both gain and loss frames may reduce framing effects (61).

Highly numerate lay people preferred risk information in numbers instead of in words alone, while those with poor numeracy skills did not (20). Slovic et al review evidence that presenting information in terms of number of individuals (e.g., 1 in 20 rather than 5%) can produce mental imagery with strong affective elements (43).

In a study that combined graphics and words, Woloshin et al asked participants to rank the likelihood of several health events, then asked them to describe each event's likelihood with words (an ordinal scale ranging from "not at all likely" to "extremely likely"), numbers (a "1-in-x chance"), and two different graphics (which will be discussed further below). They found that peoples' rankings with the verbal scale were the most reliable, usable, and strongly correlated with participants' rankings, and rankings with the "1-in-x" numbers had the worst performance (165). This result is consistent with other studies of the reliability and usability of word scales (185, 186).
5.5 Types of graphics used in risk communication

Illustrations and graphics may be used in risk descriptions, alone or combined with verbal or numeric information.

---

**Figure 5.2 (Detail)**
Some illustrations merely portray a health hazard or a symbolic warning (Figure 5.6) and are completely non-quantitative. These can be called warning graphics, and are used to convey severity rather than probability. These can be further subdivided into two types, according to the Peircean distinction between symbols that convey meaning through similarity to the concept depicted (such as the photograph at left of Figure 5.6 or, more abstractly, the skull and crossbones figure at center) and those whose correspondence to the concept is through convention, such as the radiation symbol at right (187).


### 5.5.1 Probability graphics and their properties

Probability graphics depict probabilities or numbers between 0 and 1 (or 0% and 100%); the most common formats are bar charts, icon displays, pie charts, line graphs (such as survival curves), and risk scales. They can be considered a subset of data graphics (78), which are graphics depicting quantitative data, which are themselves a subset of information graphics (78), the larger set of graphics depicting non-verbal information. Information graphics could be considered a subset of content-bearing object in SUMO, the upper-level ontology that provides an umbrella for domain-specific ontologies (141).

Graph comprehension is thought to involve a series of cognitive processes (159, 188):

- perceiving the visual features and identifying the important ones (a process dependent on how the format and design features of the graph are processed both perceptually (80, 159) and cognitively (163);
• connecting the visual features to graphical concepts such as correlation, or increase or decrease over time (a process that depends on graphical literacy and numeracy); and

• interpreting the content (a process influenced by viewers' previous understanding and expectations about the domain, as well as on supporting text and labels).

Figure 5.7: Types of warning graphics and probability graphics

<table>
<thead>
<tr>
<th>Types of Data Graphics</th>
</tr>
</thead>
<tbody>
<tr>
<td>content-bearing object (SUMO class)</td>
</tr>
<tr>
<td>information graphic (icons or combinations of icons depicting non-verbal information; includes flow-charts, diagrams, maps, charts, and tables.)</td>
</tr>
<tr>
<td>warning graphic (non-quantitative warning symbols; depictions of hazards)</td>
</tr>
<tr>
<td>data graphics (depicts quantitative and ordinal data; includes statistical graphics)</td>
</tr>
<tr>
<td>probability graphics (depicts quantitative or semi-quantitative probabilities)</td>
</tr>
</tbody>
</table>

Kosslyn (188) drew upon both the cognitive psychology and semiotics literature to produce the relatively coarse-grained ontology (background, framework, labels, and 'specifiers') for all sorts of data graphics, including graphs, flow-charts, maps, and diagrams. The effectiveness of this ontology in predicting performance on various measures of graphical comprehension has been systematically explored by (119).

A. Background

The background is the pattern or color upon which all other components are superimposed. It conveys no information but can interfere with or enhance visual perception.

B. Framework

The framework specifies the entities depicted in the display. In probability graphics, the framework consists of the x and y axes and grid of the bar chart or line graph, or the single axis of a visual analog scale. It also includes blank Likert scales and other devices meant to elicit patients' perception of risk.
C. Labels

Labels are words, numbers, or other symbols that provide interpretation of the framework (e.g., the title and axis labels) or of the data (e.g., “your risk is 10%”). The effect of numeric format on label interpretation was discussed earlier in Section 5.3. An example of the importance of labels was a study of statistical data about interactions between disease and genetics; texts were better understood and perceived as higher quality evidence than bar charts of the same data (189). However, this study should be interpreted in light of a difference between the labels on the graphs and texts that may have made the graphs harder to understand: the graphs were labeled the “relative mortality rate,” explained as “the actual number of deaths divided by the expected number of deaths,” whereas the text described one group of people as “about 20% more likely to die early deaths” than the other.

With risk graphics, symbolic labels are often used to distinguish categories such as different groups of patients depicted on the same graphic, or probabilities associated with different treatment choices (190). Information about different groups of patients or different types of diseases must be distinguished visually with colors or patterns of the elements (159), or symbols such as icons or X’s superimposed on a line graph (190).

D. Specifiers (data depiction)

The specifiers (or data depiction) consists of the line in a line graph, the bars in a bar chart, etc. The term “specifier” is adopted from (188) and (80).

![Figure 5.8: Components of graphics](image)
Component D (specifier or depiction of probability) has the following 8 properties.

D1. Level of measurement (essential property)

Probability graphics may depict risks with continuous or discrete graphic elements. (This is independent of whether the numerical label describes risk as a continuous number or a frequency.)

A. *Continuous example:* Bar chart, line graph, stacked bar chart, pie chart, risk scale

B. *Discrete example:* Stick figures, dots, squares, faces, or other icons

---

D2. Mode of depicting probability (essential property)

The specifier for probability can portray probability through different modes.

A. *Linear distance or position* (1 dimension)
   i. *continuous example* (position on a risk scale)
   ii. *discrete example* (position on scale with discrete marks)

B. *Length* (1 dimension)
   i. *continuous example* (size of bar in a bar chart)
   ii. *discrete example* (icons arranged in a row)

C. *Angle* (2 dimensions)
   i. *continuous:* slices of pie chart, angle of line graph

D. *Area* (2 dimensions)
   i. *continuous:* bubble chart, survival curve
   ii. *discrete:* icons arranged in a group or block

E. *Volume* (simulated 3 dimensions)
   i. *continuous:* boxes or cubes
   ii. *discrete:* boxes with discrete elements
F. Color or grayscale saturation or shading

i. **continuous**: color or grayscale intensity

ii. **discrete**: scattered icons in icon display

---

**D3. Proportionality** (essential property)

The mapping between the probability and the size of the graphic elements depicting it may be proportional or non-proportional.

**A. Proportional mapping**

i. **non-linear** (log scale graphics)

ii. **linear** (bar charts, pie charts, etc.)

**B. Non-proportional mapping**

i. **distorted** (mapping through a continuous but nonproportional function)

ii. **discontinuous** (mapping through a discontinuous function; breaks in the x or y axis)

iii. **semi-quantitative** (semiquantitative or ordinal scale; see Figure 5.1 above)
D4. **Part-to-whole relationship** (essential, but only for subset of probability graphics)

Probability graphics depict probability numbers, which themselves can be absolute or relative (see section 5.3 for categorization of probability numbers). For probability graphics depicting absolute risks, the part-to-whole relationship between the numerator (those affected by the hazard) and the denominator (the population at risk) may be visible or not visible.

A. **Part-to-whole relationship visible**: Example: A bar graph or stacked bar graph in which the y axis extends to 100%

B. **Part-to-whole relationship invisible**: Example: An array of icons that depicts the numerator only

![Part-to-whole visible](image1.png)

![Part-to-whole invisible](image2.png)

Fig. 5.12: Part-to-whole relationship

D5. **Scaling** (essential property)

The probability may be depicted as a proportion out of a sample of size n, or on a scale from 0 to 100%.

A. **Scaled to 100%**

B. **Scaled to n**

![Scaled to n](image3.png)

![Scaled to 100%](image4.png)

Fig. 5.13: Scaling
D6. Uncertainty (non-essential property)

Uncertainty in the form of error bars or confidence intervals may or may not be depicted.

D7. Animation (non-essential property)

Printed graphics are static of necessity. With computer graphics comes the possibility of adding animation, defined as a rapidly displayed sequence of drawings or renderings to portray processes, change, or movement.

D8. Interactivity (non-essential property)

Interactivity is the action of a user that prompts a meaningful response by the system.

5.5.2 Cognitive and perceptual effects of probability graphics

As with numbers, different graphic formats create representation effects through effects on salience, cognitive load, and demands on procedural and domain knowledge. In addition, graphic formats can create perceptual effects.

D1. Level of measurement

Continuous formats include bar charts that range from 0% to 100%, while discrete ones include icon or stick figure displays. The level of measurement of the graphic elements may be different from the level of measurement of any labels. In health risk communication, a very large amount of
research has focused on icon graphs because they appear to require fewer learned conventions (such as x and y axes).

**Level of measurement in specifiers:** In the study of willingness to pay for safer products, icon arrays and bar charts produced similar results, suggesting that in this case, level of measurement of the data depiction made no difference (29). This finding supports the suggestion made in this ontology that when icons are arranged in a block, they are judged as areas, just as bar charts are.

In a study of choices of medical treatment, an icon display reduced the influence of vivid text anecdotes (Fig. 1) (82). In this study, people were asked to imagine having angina and being offered more successful (75% success rate) but more arduous bypass surgery, or less successful (50% success rate) but less arduous balloon angioplasty. They also read anecdotes about patients who had had the procedures. The number of anecdotes describing success strongly affected participants' choices. When the proportion of successes in the anecdotes was the same as the treatments' success rates (for example, when 3 of the 4 bypass stories described a treatment success), respondents became more likely to choose the more successful alternative (bypass). When one anecdote described success and one a failure, most respondents chose the less arduous treatment (angioplasty). The anecdote effect was significantly smaller when respondents saw icon displays depicting the two treatments' success rates (82). The icon array showed the part-to-whole relationship and the square icons were touching, so the display might have been visually processed as areas rather than as discrete icons.

A qualitative study of 40 women found that simple bar charts depicting absolute lifetime risk of various events were preferred over line graphs, thermometer graphs, icon arrays, and survival curves (173). Participants wanted graphics to be supplemented with text.

In a focus group of women, participants preferred icon arrays with smaller denominators because they seemed simpler but also tended to think that graphics with larger denominators portrayed risks as smaller (81). The findings are not consistent with the common ratio-bias effect, in which risks described as ratios of small numbers are considered smaller than numerically equivalent risks described with large numbers (e.g., 1 in 20 is considered less likely than 1 in
In another focus group study with low-income women, participants preferred seeing an individualized risk estimate depicted as a bar chart with an ordinal scale (low, average, or high risk) rather than as an icon array or a percentage, and rather than a bar chart showing a series of relative risks for women in different risk categories (86).

Fuller et al used several tasks to assess how elderly patients interpreted discrete icon displays (191). The patients could match percentages to icon arrays displaying different proportions (70% to 98% accuracy for different tasks). They were less accurate when marking the graph to show probabilities (either ratios with different denominators [38% to 79% accuracy] or percentages [51% to 98% accuracy]). The authors did not assess whether the graphs were successful in conveying the personal applicability of the risk. A short report by the authors (8) described similar results but with few details.

**Level of measurement in framework:** Women estimating their risk of breast cancer provided different estimates when asked to show their own risk on an icon display (a grid of 100 female figures) and a continuous scale (a horizontal line anchored at 0% and 100%) (166). Icons elicited risk estimates that were higher and farther from epidemiological risk.

Woloshin et al asked participants to rank the likelihood of several health events, then asked them to describe each event’s likelihood with words (an ordinal scale ranging from “not at all likely” to “extremely likely”), numbers (a “1-in-x chance”), and two types of horizontal risk scales (165). Rankings with the verbal scale were the most reliable, usable, and strongly correlated with participants’ rankings, and rankings with the “1-in-x” numbers had the worst performance.

Licensed anglers were shown risk ladders describing the hazards of eating contaminated fish in discrete numbers or ordinal categories (“higher risk”, “moderate risk”, and “lower risk”) (192); 57% preferred the quantitative ladder.

**Human figures versus other icons** Although Stone et al found no differences in behavior with asterisk and face displays (28), Schapira et al’s qualitative study found that women considered human figure icons (like those in Fig. 2) to be more meaningful, easier to understand, and easier to identify with than bar charts (81). Women in one of the focus groups, which had a lower mean age and educational level, perceived risk of breast cancer as larger when it was
shown as a part-to-whole human icon display than when it was shown as a part-to-whole bar
graph; however, no quantitative results were collected (81). Some participants said the icon
display suggested population risk, while a continuous scale suggested personal risk. In direct
contrast, Royak-Schaler et al found that focus groups of low-income women preferred a part-to-
whole bar chart to a part-to-whole icon array; however, in this study the bar chart had evaluative
labels ('high risk,' 'low risk', or 'average risk'), but the icon array did not (86).

D2. Mode of depicting probability

Research in psychophysics and graphical perception has been applied to determine the visual
displays most likely to be judged accurately. The classic psychophysical research of Cleveland
and McGill ranked the "specifiers" of data by the accuracy of judgments (79, 80). In order of
accuracy, these specifiers were:

- positions against a common scale in 1 dimension (such as positions on a risk scale);
- positions against non-aligned scales (such as positions on separate scales);
- lengths (such as bars in a bar graph); angles (e.g., slopes of a line graph or slices in a pie
  chart);
- areas (such as squares or circles);
- volumes (cubes);
- color or gray-scale shading or saturation.

A number of experiments have confirmed the importance of these basic graphic perception
findings, although some have challenged a few of the rankings (80, 163, 193). Carswell's
metaanalysis (119) determined that this Cleveland and McGill taxonomy (or "basic tasks model")
performed better than Tufte's 'data-ink ratio' (78) in explaining several hundred studies of
performance and judgment with graphs. The basic tasks model was better at predicting
performance in point-reading tasks (such as "What is the height of the bar in the bar chart?"?) and
local comparisons ("Which bar is higher? how much higher?") than in integrative tasks (such as
"Is the variability depicted here large or small?"). The meta-analysis also found that the model
was somewhat less successful at predicting performance on contrasts between graphics in the
top ranks (119), which may suggest that differences between these ranks may be too small to be picked up by most studies. In addition, Hollands conducted a series of studies in which estimates of proportion with angles in a pie charts were moderately better than with part-to-whole bar charts, and much better than with side-by-side bar charts (163). This suggests that (a) part-to-whole relationships have a strong effect, and (b) the relatively low rank for angles may need to be raised.

Feldman-Stewart et al assessed speed and accuracy of students' and patients' judgments with 6 formats: vertical part-to-whole bar chart, horizontal part-to-whole bar chart, pair of numbers, part-to-whole icon graph with random arrangement, icon graph with the icons arranged in a block, and pie chart (162). Participants were slowest and least accurate at judging the larger of two quantities with the pie chart and the random-arrangement icons. Estimates of the differences between quantities were best with number pairs and sequentially arranged icons. Participants performed no better with their preferred formats (162). These findings support the suggestion that random icons may be judged by color saturation, the least accurate of the Cleveland-McGill specifiers.

One trial compared graphics for conveying risks to physicians (30). Physicians saw data from a fictitious clinical trial in which one treatment had a high failure rate. Clinical trials may be halted midcourse if results in one group are much worse than in the other; the physicians were asked if the data warranted halting the trial. Five formats were given (tables of success rates, tables of failure rates, pie charts, stacked bar charts, and icon arrays). Most noticed the high failure rate in icon arrays and deals; fewer did with pie charts or stacked bar graphs (30). However, most liked the bar graphs and disliked the icon array. These results are consistent with research showing that proportions are difficult to judge when mental summation is required (194) although not with a finding that pie charts were superior when mental summation of slices was required (193). The authors suggest that the icons' success was due to the framing effect of drawing attention to the failures, which would be consistent with the Stone et al 'foreground effect' (27). Another explanation is that the discrete icons could be counted, but the other graphs required area
estimation; however, earlier findings that results with bar charts were the same as results with icon arrays (28) argues that viewers may not be using counting as a strategy.

**Linear distance and lengths (1 dimension):** A risk table, ladder, or scale depicts a range of risks from very low to very high as context for an individual risk. When risks are ranked vertically in a table, a graphic is called a risk ladder, and when they are horizontal, the graphic is generally called a risk scale or visual analog scale. Because position on a risk ladder or scale is evaluated as distance from a baseline, Lipkus and Hollands (24) propose that they exploit the most efficient of the Cleveland and McGill basic (80) visual perception skills.

More systematic explorations of risk ladder design have been done in environmental health. When a person’s exposure to an environmental hazard was explained by referring to a location on a risk ladder, perceived risk was associated with location on the ladder rather than numerical magnitude of the risk (195). The ladders also illustrated unfamiliar concepts with text and graphics, such as icon arrays of the number of cigarettes needed to produce a cancer risk comparable to a given level of radon risk (195). Johnson and Slovic compared numbers and a risk ladder for communicating uncertainty (confidence intervals) about risk estimates (196). The numbers were ratios with different denominators, and the graphic showed no part-to-whole information. When compared to numbers alone, the ladder did not affect perceived risk; it did decrease trust in the information but also improved ability to notice the full range of possible risks (196).

**Areas and angles (2 dimensions):** In the Stone, Fagerlin, and Schapira studies, the icons depicting people affected by disease were arranged in blocks, so risks could be estimated by judging the length or area of the block, two tasks that exploit relatively strong visual perception skills. The Royak-Schaler experiment used figures arranged in a block, but they were staggered rather than in neat rows, which could have made it more difficult to use visual area judgment.

Survival and mortality trends are cognitively complex because they involve changes over time. When survival or mortality data are presented graphically, changes in risk are inferred from curve slopes or areas (78, 79). Part-to-whole relationships are available, though not very salient, when the y axis extends to 100%. Because line graphs portray data points as a single visual
element, viewers do not need to integrate the information themselves; line graphs thus help experts perform complex tasks such as assessing rates of change (108). However, Armstrong et al showed that only 74% of a sample recruited from a jury pool could interpret a survival curve well enough to determine the number of survivors at various time points, and only 55% could calculate the difference in survival between two time points (25). After a training exercise, ability to interpret survival at one time point improved but accuracy in calculating differences was unchanged. The effect of learning can also be inferred from older studies in which choice of treatments was strongly affected by the amount of instruction in interpreting survival curves (197). With minimal explanation, patients tended to choose the treatment with better long-term survival; patients given extensive explanations were more likely to take medium-range outcomes into account (197). Physicians were more likely to be influenced by middle parts of the curves than were patients (198, 199), which could also be due to education. Survival curves may reduce a tendency to overweight immediate survival by drawing attention to longer-term outcomes. When patients were given a choice between treatments described with survival percentages, 59% chose the treatment with better immediate survival; this dropped to 34% when they viewed a pair of survival curves (200). The order in which patients viewed survival graphs significantly affected their preference for short-term versus long-term survival (201, 202). More educated patients were less likely to choose the treatment with better short-term survival (201). Participants answered comprehension questions more accurately with survival curves or both survival and mortality curves than with mortality curves alone; the effect was strongest in the lowest educational group and among non-whites (61). In this study, participants were asked to imagine being at high risk for colon cancer and were given a choice between colectomy and an easier but less successful alternative (annual exam). They were less likely to choose colectomy when viewing mortality curves; the effect might have been due to the reduced understanding of information with mortality curves (61).

In another study, people were influenced more by the distance between curves than the numerical differences (Fig. 7) (181). If a pair of 15-year survival curves are displayed on an x-axis of a certain length, they will diverge more than if the first 5 years of data are stretched and
displayed on an axis of the same length (203). This flattening effect markedly reduces the
difference between peoples' estimates of treatment effectiveness (181).

In a study of women with BRCA1/2 mutations who were deciding on possible prophylactic
measures against breast cancer, those who received a set of personalized survival curves were
more satisfied with their decisions than those who received a similar educational booklet without
survival curves (62). However, the survival curves did not change their actual decisions.

Volume (3 dimensions): Simulated volumes are relatively rarely used in risk graphics, but
some examples are available. Presenting data in 3 dimensions encourages meaningful
interpretation of relationships among all three variables, although it also reduces the accuracy of
reading individual data points (159), which consistent with volume judgment (80). Adding an
extraneous third dimension to bar graphs somewhat reduces readers' accuracy in judging the
height of bars (159, 204). However, this effect may not be very serious. For example,
manipulating the height of a comparison bar in the bar chart produced a much larger decrease in
accuracy than did the extraneous depth cue (159, 204).

Color or grayscale saturation: Risk is sometimes portrayed with icon arrays with icons
scattered throughout the field, instead of grouped in a line or block.

With this type of random arrangement, area judgment is not available to the viewer. It is
possible that comparisons are made through a judgment of gray-scale intensity or saturation, one
of the least accurate of the Cleveland and McGill visual perception tasks. Accordingly, I have
proposed "gray-scale saturation" as the final mode of depicting probability. However, it is possible
that some other process is occurring, such as mental summation of small areas (also a relatively
inaccurate task according to (194). Alternately, viewers might be counting icons. Additional study
might help to clarify how people are making these judgments.

In genetic risk counseling, such a random arrangement has been described as helpful in
promoting understanding of chance (205). However, viewers were less accurate at estimating
proportions in random arrays than in sequential ones (162). In the Schapira et al study, women
disliked the random arrangement because they said they could determine the probability only by
counting (81). However, some in that study said it better conveyed the idea of randomness.
D3. Proportionality

In proportional graphics, the size of the graphic element is proportional to the quantity it depicts. For example, in linear proportional graphics, a segment 10 units high depicts a risk exactly twice as large as one 5 units high. This property is considered desirable for data integrity (78, 167). Distorted graphics such as tilted pie charts may inflate the apparent size of particular risks.

One risk scale, the Paling Perspective Scale (110, 206) depicts a range of probabilities on a logarithmic scale from 1 in 1 (certainty) to 1 in 1 trillion, centered on 1 in a million, described as “effective zero.” This horizontal scale was used to compare the risks from a blood transfusion (such as contracting HIV) with other hazards such as the annual chance of dying in a car accident. The authors have not published any data about viewers’ interpretations of the log scale. The ladder was as effective as numbers alone in increasing knowledge and reducing dread about rare hazards of transfusion, but was not tested against other visual displays.

The risk magnifier study described earlier explored the usefulness of a discontinuous framework, a horizontal risk scale ranging from 0 in 100 to 100 in 100, supplemented with an image of a magnifying glass to illustrate probabilities smaller than 1% (165). The magnifier scale permitted people to make lower estimates for very rare events. A subsequent study by another research group compared the magnifier with the standard horizontal risk scale (207). This work confirmed that the magnifier scale enabled appropriately low estimates for very rare events but also showed that it substantially lowered risk estimates for more common events (207). This effect was seen when participants estimated risks of various health events without being given numeric information about the magnitudes of those risks.

D4. Part-to-whole relationship

When the part-to-whole relationship is visible (e.g., with a pie chart), viewers can generally use rapid, automatic perceptual skills to estimate what proportion is represented by the pie slice (163, 194). However, when the part-to-whole relationship is not visible, the proportions are not visible and must be inferred. In other types of graphics, e.g., a bar chart in which three bars sum to
100%, the proportions may have to be estimated by visually summing the bars (194), a more cognitively effortful and consequently less accurate task.

The importance of the part-to-whole relationship on behavioral outcomes is supported by a series of studies by Stone and colleagues (27, 29) to follow up a 1997 study (28). In these studies, undergraduates received information about pairs of fictitious products, each carrying a small probability of a harmful effect (e.g., tire blowouts with tires). Participants estimated how much the safer product would be worth. The graph showed the number harmed but the at-risk group was provided as a number, so the part-to-whole relationship was not available visually (Fig. 2). People were willing to pay more (i.e., were more risk-averse) when the number harmed was depicted as an array of stick figures, asterisks, or faces, or as a bar graph than when it was number. In one study,(27) the graphs were compared to graphs that did portray the part-to-whole relationships visually, for example, a bar graph showing only the number affected compared to a stacked bar graph of those affected as a proportion of the entire group. Participants were willing to pay more for the safer product with a graph that did not show the part-to-whole relationship than when given a number. However, when they saw part-to-whole relationship in the graph, they were not willing to pay more than when they saw numerical probabilities. Stone et al suggest that risk aversion for rare events is the result of graphs that fail to show part-to-whole relationships, not of all graphs, and label it ‘foreground effect’ (27, 29).

These studies provide a new perspective on an older pair of studies that used dots to depict only the denominator of the probability; viewers were told that the risk was 1 against the number of dots. An icon display of the risk of rare side-effects from a vaccine increased the number of subjects who said they would get vaccinated, presumably by focusing their attention on the denominator (208). However, a similar study could not replicate the effect (85).

**D5. Scaling**

Judgments of proportion may be complicated by scaling issues. With graphics such as bar charts, a probability may be depicted graphically as a number out of N people, or as a percent out of 100%. With the former, different bars will have different sizes, making it difficult to compare
proportions in different bars. That is, it is more difficult to see that roughly the same proportion of each bar is black in the left-hand version than in the right hand version (194).

![Fig. 5.15: Scaling and proportion](image)

However, pie slices are relatively easy to compare regardless of scaling, because viewers apparently judge the angles rather than the areas (163).

**D6. Uncertainty**

In scientific text, particularly in the discipline of epidemiology, risks are frequently bracketed by a standard error, confidence band, or confidence interval. Graphically, in scientific publications, these measures of uncertainty or precision are commonly depicted as T-shaped error bars (on bar charts, dot plots, or line graphs) or confidence bands (on survival curves).

However, only rarely is this uncertainty about the risk communicated to the public. This is probably because a number of studies have confirmed that most lay people are both unfamiliar with and uncomfortable with the concept of uncertainty in science. In one series of questionnaire and qualitative studies, lay people were given news stories about health risks portrayed either without uncertainty (e.g., a single probability) or with uncertainty (e.g., a range of probabilities) (196). Although most people were not familiar with the concept of uncertainty, simple presentations including graphics did seem to explain the idea to many people. However, the presentations of uncertainty were interpreted in two opposite ways as honesty or incompetence. General risk attitudes and perceptions had a stronger effect on interpretations than did the uncertainty manipulation.

Genomic Health’s report for Oncotype DX depicts the mean estimated cancer recurrence rate for all women as a curved line, accompanied by uncertainty depicted as a confidence band of
The explosion in computer-mediated education and communication has made animation widely available. An animation is a rapidly displayed sequence of drawn or rendered images, portraying processes involving change or motion. Animation processes include appearance/disappearance, movement, and change in shape or color. Animations are now commonplace on television (e.g., weather maps) and on informational websites such as Wikipedia and WebMD.

At their most trivial, animations include any movement or change imposed on text or images, such as flashing, color changes, or progression of words across a screen. However, when the processes depicted in the animation constitute a model of a system, the animation is a simulation. Examples include depictions of how to use an asthma nebulizer in (88) and animated cellular processes in diabetes (http://diabetes.webmd.com/healthtool-type-2-animated-guide). According to computer games theorist Ian Bogost, such techniques make possible an alternative to traditional verbal rhetoric (communicating through words) and visual rhetoric (communicating with images). This alternative is procedural rhetoric: the use of procedures to communicate information about procedures (112).

Research on the use of animated simulations for education has had mixed results. Animations attract attention by increasing salience (209-211). They can make health-related materials more interesting and increase consumers' motivation to use them (102, 212, 213). It might seem obvious that since static graphics are useful in depicting spatial dimensions, animated graphics should be successful at depicting change over time. In fact, Bogost argues that "procedural rhetoric," or the use of procedures to communicate information about procedures, has joined verbal and visual rhetoric as communication modalities (112). Procedural animations, however, are not necessarily the best way to illustrate procedures. Many animations have been found to be confusing, and animated explanatory materials generally fail to produce better learning than static materials (102, 159, 213-217). It appears that animation itself can
sometimes interfere with learning and understanding. In many studies of animated informational materials, the animation appears to increase cognitive burden by providing too many details, requiring the viewer to hold past states in memory, and failing to visually distinguish relevant from irrelevant information (95, 99, 100, 113, 217). Tversky has further shown that in some cases, the use of animation does not match viewers' pre-existing mental representations of the process: an example is map directions, which most people conceptualize as a sequence of discrete steps rather than as a single continuous process (95, 113, 214).

**D8. Interaction**

Interactivity involves the user taking some action, and the system supplying a meaningful response in return. Several researchers distinguish between simple interactivity (the ability to start, stop, or repeat an animation, as with DVDs or YouTube videos) and feedback or loop interactivity (in which the user input changes the state of the system, and these changes constitute meaningful information) (128, 218). Examples of feedback-type interaction include: inputting personal health risk factors into an on-line risk calculator and receiving a personalized estimate (124); playing electronic games (87-89, 91, 92, 112); gaming or interacting in online multiuser environments (219); and using training simulators (130, 131). Procedural skills learned through virtual experience are often transferable to the world (130).

Learning and instruction research suggests that adding interactivity to animations resolves many of the problems with animations discussed in the previous section. Simple interaction, such as stopping or replaying an animation, can reduce cognitive load by allowing the viewer to chunk information and examine past system states at leisure. Loop-type interaction allows users to meaningfully explore the system to build a more realistic mental model (99, 100, 135). In building a mental model of a system, one particularly useful task enabled by interactivity is the iterative exercise of predicting system output, and then determining whether the prediction was correct (95).

Some previous applications of interactivity to teach health information have been automated quizzes and flash-card style drill learning, with animated cartoons that serve primarily to increase
interest (for example, see http://asthma.starlightprograms.org/ and (93, 220). Others have focused on interactive exchanges of verbal and numeric information. Emmons and colleagues developed a computer tool that allowed participants to input their own risk factors and receive a number representing their risk of colorectal cancer (124). In this study, patients using an active engagement version that allowed them to see the effects of altering their own modifiable risk factors were more accurate in their risk perceptions than patients using a passive version, although the effect was not statistically significant (124). However, this tool allowed users to exchange only verbal and number information.

By contrast, others have used interactivity to provide a rich virtual experience. A series of studies of choice behavior used interactions to provide experiential learning about the chance and size of monetary payoffs (31, 32). In these studies, experimenters assembled two decks of real or virtual cards, each displaying the amount of money to be won by drawing that card. Participants could sample as many cards from the two decks as desired, then choose a deck to draw a card for a real monetary payoff. Decisions in this "learning from experience" paradigm were quite different from those made when risks were described in numbers or graphs. For example, as described in prospect theory (11), people usually overweight the probability of rare events (for example, avoiding gambles that carry a small chance of losing a large sum, even if on average, such a gamble would be a good bet). However, participants underweighted the probability of rare events when learning from experience (31, 32).

Similarly, a number of other health-related games and systems provide interactive simulations on the basis of fairly realistic models of genuine physiological systems, and thus provide experiential learning about the health condition (88, 91, 92, 122). For example, in one of several asthma games developed for children, exposing the game character to allergens will produce asthma attacks in the character (88). Extended use of these asthma games has been shown to improve both knowledge and measurable medical outcomes among children (88, 91, 92). One on-line diabetes simulator allows diabetes patients to input their food intake and other personal data and then see personalized glucose estimates; in a small trial, patients who used the system had in lower hemoglobin A1c levels and fewer episodes of hypoglycemia than did
patients participating in diabetes education seminars (122). Even in several of the Grand Theft Auto games, feeding a character fast food (and neglecting to visit the virtual gym) fattens the character; also, a fat character sometimes elicits derogatory remarks from peripheral non-player characters, showing that the game includes a simplified social model as well as the physiological one (112, 221). The Serious Games Initiative http://www.seriousgames.org and the affiliated Games for Health http://www.gamesforhealth.org, sponsored by the Woodrow Wilson International Center for Scholars, are attempts to stimulate game development.

The advantages of interactivity are consistent with the constructivist learning model. The central idea in constructivist learning is that learners actively construct new knowledge on the basis of their previous knowledge. Meaningful learning comes from activities that encourage active knowledge construction and engage previous knowledge, rather than from the passive receipt of information (120). In addition, according to the elaboration likelihood model, any attitude changes deriving from such active information processing will be more stable and persistent than changes deriving from peripherally processed information (66). In one health-related example, Natter and Berry invited consumers to draw graphs of their own health risks, as well as to write reflective essays (67). Consumers in the active learning condition understood the risks better and expressed more satisfaction with the information than people in the passive condition.

5.6 Summary and implications

The research on risk graphics and risk numbers is extensive. In particular, many researchers have manipulated single features of the graphics (e.g., part-to-whole relationship, or proportionality) and assessed effects in context of one or more of the risk communication purposes (e.g., information, decision support, or persuasion). Particularly well studied are the "modes" of presentation (length, angle, area, etc.), which have been arranged in a well-substantiated hierarchy in terms of their effect on accuracy of judgment. In addition, icon or stick figure graphics are relatively well-studied, with a consensus being that they are perceived as

---

1As mentioned, Peirce distinguished between symbols that convey meaning through similarity and those that convey meaning through convention; he also proposed a third type that convey meaning through causation, e.g., a weathercock that points the wind direction (186). Interactive systems could fall in this category.
useful by a variety of people including those with low literacy because they require little knowledge about graphical conventions; that they improve basic comprehension and knowledge; and that they can be exploited for either persuasion (such as drawing attention to the benefit) or neutral decision-making (such as mitigating unwelcome decision effects from vivid verbal anecdotes).

The body of work discussed here is sufficient to construct an ontology of the features of graphics that affect cognition, perception, or behavior. It also helps to resolve some apparent contradictions in the literature by showing that the contradictory results are apparently attributable to a feature not considered by the researchers (such as the failure of a bar chart because of poor axis labels). This review also clarifies that different studies have pursued different goals, with much of the psychological literature focusing on the effects of graphics on comprehension or knowledge, educational literature focusing on effects on learning, medical and decision science literature on neutral decision support, and public health and preventive medicine literature tending to focus on effects on persuasion. The resulting ontology could be the basis for a decision support system to help guide the best choice of graphics for different communication purposes.

However, the review shows that some features of risk graphics have hardly been studied at all, or the existing studies have not identified effective presentation methods. Depicting scientific uncertainty (i.e., standard errors or confidence intervals) is an unsolved problem. As mentioned above, laypeople tend to be unfamiliar with the concept of scientific uncertainty and uncomfortable when experts portray information as uncertain. Some graphics (as the Oncotype DX report shown earlier) display confidence bands or intervals, but the effect on patients is unknown. Acknowledging and showing the range of scientific uncertainty would seem to be particularly ethically important for patients making decisions, suggesting that further research is warranted. Another set of single-feature studies are possible with stick figure graphics. For example, the Cleveland and McGill mode literature and the Hollands summation model would suggest that proportions depicted in randomly arranged stick figure graphics will be estimated with less accuracy than sequential arrangements. If this is found to be true, however, it is an open question how this result would affect learning, persuasion, or decision support.
Two important features that have been insufficiently researched are animation and interactivity, now easily available on any computer. Much of the existing research in these domains comes from cognitive or educational psychologists working on relatively general topics (such as navigating maps, following assembly instructions, or learning a mechanical concept). These findings strongly suggest that adding interactive animations will increase interest in the information and support in-depth learning. However, studying how best to adapt these findings for health and medical purposes will require more input from researchers in health-related fields. In particular, relatively little work has been done on applications of animation and interactivity to neutral decision support or persuasion. Several online epidemics in multiuser game environments suggest that these environments provide rich opportunities for learning about probability. In one example, educators on a student-oriented multiuser environment created a disease that was transmitted from avatar to avatar (219). Such virtual epidemics could be used to provide participants with an experience of the probability of disease. For example, contacting an infected character could result in disease with a biologically plausible probability, instead of with certainty. To provide a wider perspective, players could be given views of disease spread among entire societies. Giving the users the ability to change the transmission probabilities or other features of the algorithm, and then view the results, could create powerful educational experiences about not only probability but also more complex epidemiological concepts such as herd immunity.

More importantly, the review shows that the risk graphics literature is weak on integration. For example, although one set of researchers has shown that a graphic that suppresses the part-to-whole relationship can be persuasive, and a second set of researchers has shown that replacing graphics with vivid verbal anecdotes can also be persuasive, it is not known which of these efforts is more persuasive. A standard evaluation method would be needed to compare different effect sizes and identify the manipulations with the highest efficacy. Similarly, the single-feature studies do not elucidate the optimal combinations of features for different risk communication purposes. Studies of multi-feature manipulations would be useful in this regard. An even larger integration issue is how best to combine graphics with verbal text and numbers for different purposes. Open questions include: When persuading people to adopt a health precaution, is it better to use
graphics to illustrate the probability of the health risk or its severity? Is it better to use graphics to
illustrate the health risk, or to illustrate the benefit associated with the precaution? or some
combination of these? For educational purposes, which information should be conveyed with
words, which with numbers, and which with illustrations? Should graphics be accompanied by
explanations to improve comprehension, and if so, what types of explanations?

Finally, this review suggests that an important under-researched topic is the interaction with
the viewer's cognitive competencies such as numeracy. Numeracy strongly affects viewers'
comprehension of numerical information, as well as trust in and comfort with that information.
There is sufficient evidence to hypothesize that people with poor numeracy will want graphic
illustrations to help interpret risk information, but also that they may be more strongly led astray
by less relevant features of the graphics. For example, Peters found that people with poor
numeracy tended to be more strongly influenced by the number of black balls in an urn than by
the proportion of black balls, which could lead them to make poor choices. There is also sufficient
evidence to hypothesize that explanations or training will be needed to help people with poor
numeracy interpret complex graphics such as survival curves. Subsequently, additional research
could focus on applying the findings to tailoring. Although there is extensive research on tailoring
text to low literacy levels, there is little research about the best ways to tailor graphics and
numbers (especially risk numbers) for lower levels of numeracy.

As I have suggested earlier (in Chapter 2), communication about health and science issues
should be considered not a one-way transmission from the expert information provider to the
patient, but a three-way interaction between the patient, the expert provider of information, and
the information artifact (such as the text or the graphic). Studying all three participants and the
interactions between them is the best way to ensure comprehension of information for various
risk communication purposes. Such a research program requires combining perspectives from
health communication and health literacy, cognitive psychology and decision science, educational
technology, clinical medicine, informatics, and public health. Resolving these issues would
provide the specific detail needed to use the ontology presented in this chapter as the basis for
decision support logic to help communicators match appropriate risk descriptions and graphics to specified communication goals.
CHAPTER 6. RESULTS OF QUALITATIVE STUDY


Five focus groups were held between October 2006 and March 2007. Group size ranged from 3 to 5, except in one group conducted during a heavy rainstorm when only 2 participants arrived; I proceeded because they were eager to continue and did not want to reschedule. The total number of participants was 16.

6.1 Themes

Several major themes arose in the qualitative analysis. In describing them, I have attempted to represent the differing opinions and perceptions throughout the discussions.

- Emotional impact: The coding for emotional responses showed that the interactive program was associated with more expressions of emotion by the participants than the other graphics and visuals (for example, 17 comments were coded as emotional responses for the interactive program, compared to 4 for the bar chart). Several participants, when playing with the computer program, expressed dismay when a stick figure icon “got the disease.” One woman said she didn’t want to “play” anymore. Conversely, a participant who was exploring a low risk said she was relieved because she clicked so many times without finding a figure with the disease. Some people expressed impatience that the interaction took too long, delaying their arrival at the page with risk-reducing advice. The participants rarely expressed any emotional reaction to the printed illustrations of stick figures or to the bar chart, and only one person expressed an emotional reaction to a percentage (she cheered when she saw that her 10-year risk of heart disease was “less than 1%).

- Impersonal numbers: Participants tended to describe quantitative information as impersonal and irrelevant, saying for example, “It’s not warm,” and “Percentages always have to do with other people.” One said that if she had a doctor who told her nothing but percentages, “it’s going to make me want to go online or go make another appointment with a doctor who can make
it clearer." However, another said that personalized risk information in the form of numbers "says a lot to me."

When invited to describe a percentage risk to another person such as one of the characters in the scenarios, most participants suggested describing them in terms of numbers of people even before seeing any of the stick figure illustrations.

- **Stick figures:** Most participants preferred the matrix of stick figures to the bar chart because it was "clearer that you're talking about human beings and not statistics." However, other participants felt that the stick figure display was somewhat overwhelming and "a lot to look at," and that the bar charts were "straightforward." Most thought the random arrangement (Figure 1A) was more difficult to count than the sequential arrangement (Figure 1B). However, positive comments about the random arrangement outnumbered negative ones (8 positive to 2 negative). Many described it as realistic, e.g., "the more realistic way to depict the chance." One said that the sequential arrangement "can give a person a false reading" because it made it too easy for her to imagine that she would *not* be affected by the risk.

- **Interaction:** The interactive process of clicking squares to see the icons underneath was interesting to some of the participants. One said, "It's like a game because you're playing around with it. That's what I like about it. Because you learn too." However, some others were impatient with the interaction, either because they said they already understood the percentage, or because the module emphasized the negative (getting the disease) too strongly. Most said they enjoyed the interactive aspect of being able to input personal information and get tailored output.

- **Physician-patient communication problems:** Many of the passages in which participants mentioned their physician or their relationship with their physician were coded with negative emotions and with a code indicating that they felt a need for more information. In particular, participants offered stories in which they had received information from a doctor and not fully understood it, as depicted in the scenarios. One woman recalled having a serious medical condition that had not followed the disease course she had expected, and said, "Instead of me pushing further with the doctor in conversation, I decided to seek information on my own." She also said: "You don't want to seem stupid so you just don't ask." Another said: "I feel like
sometimes the doctors are intimidating... they kind of rush you.” Participants said they would go
to other sources to find more information, most often mentioning the Internet and the library.
There were some expressions of distrust of both medicine and doctors, e.g., suggestions that
prescription drugs were riskier than herbal preparations, and that doctors avoided explaining the
risk of drug side effects for fear of being held liable for those side effects.

- **Comparisons:** Participants using the interactive tool did not suggest adding interpretive
information such as “this is a high risk,” or comparative information such as “this risk is higher
than average.”

### 6.2 Formative development

Participant suggestions and observations of participant behavior were used to update and
develop the software prototypes throughout the project. Most participants found the prototype
easily usable, although one who had virtually no experience with computers found it intimidating
and declined to try using the computer at all.

Examples of changes to the prototype include:

- **Stick figure arrangement:** All groups saw both randomly arranged and sequentially
arranged stick figure illustrations. As described, participants generally agreed that it was easier to
judge the proportion in the sequential arrangement of stick figures (Figure 1B) but described the
random arrangement as more “realistic.” The final version displayed the sequential arrangement
first (Figure 1B) so that viewers could easily see the proportion. When viewers advanced to the
next screen in the module, they saw the random arrangement of the same quantity (Figure 1A).

- **Repeated sampling:** The first two focus groups saw an early prototype allowing users to
click multiple times to see multiple versions of Figure 1A showing different random arrangements,
in an interaction designed to simulate the process of drawing different samples from a population.
In general, however, participants said they did not understand this interaction or like it. For the
third group, I replaced the repeated sampling (which could be repeated as long as the user
wanted) with a sequence of two screens showing different randomly arranged “samples.”
However, many of the participants still did not find this very useful. In the end this interaction was removed.

- **Action steps:** In early groups, I described possible steps people could take to reduce their risk as part of the scenarios presented orally. However, the oral descriptions did not appear to be as salient to the participants as the visual materials. Several participants criticized the doctor characters in the scenarios for not describing preventive actions in sufficient detail. I therefore integrated the “action steps” into a page appearing after the final graphics for the final focus group.

- **Interaction:** The prototype for the first three groups showed a grid of squares with one of them colored blue. The user was invited to imagine that he or she was standing underneath it; clicking that square revealed all the stick figures including the one representing "you." The user could then reset the grid of squares and try again. As mentioned, this sometimes elicited a strong emotional response. I was concerned that if the image was too threatening, people might refuse to use the program or reject its message (3). To attenuate the impact somewhat, I changed the wording after the third focus group so that the illustration was described as showing someone among “a group of people” rather than "you." This appeared to reduce the negative emotional responses in later groups without eliminating them. After reviewing the final focus group transcript, I also developed a modification (Figure 6.1 below; subsequently used in the questionnaire study: Chapter 7) that would allow users to click on multiple squares, rather than only a single square. The goal was to satisfy participants’ desire to reach the “action steps” advice more quickly while still encouraging them to interact with the graphic.

- **Text:** Participants frequently critiqued and offered suggestions about the explanatory text on each page of the interactive program. For example, the early versions explained the percentage risk as “6 in every 100 people,” but the subsequent illustration showed 1000 stick figures. Although some participants thought that the phrase “every 100” was clear enough, others found it confusing. I expanded the text after the second focus group to read: “A 6% chance means in the past, when doctors examined 100 people like you, about 6 of them had a serious heart attack in the next 10 years. And among 1000 people, about 60 had a serious heart attack.”
Figure 6.1: Interactive search graphic
(Working example at www.dbmi.columbia.edu/~isa7002/dissertation/final/Search.php; requires Adobe Flash Player; graphic was displayed at 350 x 320 pixels or 4.7 x 4.3 inches)

A. The first view is a grid of squares
B. Clicking on a square shows the figure under it

C. Yellow figures are unaffected
D. Blue figures have the disease

E. Finding a blue figure triggers a cascade that reveals all the figures
CHAPTER 7. RESULTS OF QUESTIONNAIRE STUDY

The questionnaire was subjected to preliminary analysis for scale reliability (Section 7.1) and then was administered to 165 subjects. Results are presented in Sections 7.2 through 7.4.

7.1 Test-retest reliability

Test-retest reliability was assessed by having 9 participants (not involved in the qualitative study or the pilot testing) take the questionnaire twice over an interval of 2 to 3.5 weeks (median: 18 days). Reliability was good for all subsections (Table 7.1).

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Average test-retest item correlation*</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opinions (2 items)</td>
<td>0.89</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Estimates (6 items)</td>
<td>0.91</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>STOFHLA (items completed out of 36)</td>
<td>0.99</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Numeracy (8 items, coded as correct/incorrect)</td>
<td>0.83</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Perceptions/decisions (20 items)*</td>
<td>0.66</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Personal risk estimates (2 items)</td>
<td>0.75</td>
<td>.002</td>
</tr>
<tr>
<td>Demographics (6 items)</td>
<td>0.98</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

*Spearman’s ρ for categorical variables (e.g., opinion scales); Pearson’s r for continuous variables (e.g., estimates); tetrachoric correlations for binary variables (numeracy scale).

Two subjects were excluded from analysis of this subscale because they failed to follow instructions and signed into the wrong arm of the study for their retests.

The correlations for the total score for the STOHLA (r = 0.68; p = .04) and the numeracy scale (rho = 0.69; p = .04) were also good. In the STOFHLA, 2 people timed out on the scale the first time but not the second, which created the appearance of a learning effect. As both were college graduates, it seemed possible that the time-outs resulted from distractions rather than low literacy. As a result, I changed the instructions to stress the importance of working straight through this section without taking a break (although the time limit was not disclosed). The changes in the numeracy scale did not appear to be due to learning effects; 2 scores improved, 2 worsened, and 5 remained the same. The slightly lower reliability for the perceptions/decisions scale was not unexpected: one hypothesis had been that the interaction with the graphic would alter risk perceptions and decisions, and preliminary data did show a correlation between click number and risk perception (discussed below).
7.2 Recruitment and demographics

Data were collected from the Virtual Lab pool in July 2008. A total of 101 people responded with complete data, but one was dropped from the analysis for completing the questions in an obviously rushed (<6 minutes total) and nonsensical (score of 5 out of 36 on the TOHFLA and 1 out of 8 on numeracy) way. In addition, 65 participants were recruited from Columbia Presbyterian Hospital waiting areas (the pediatric emergency department, the pediatric asthma recovery room, and the dental teaching clinic) from July through October 2008.

Only complete questionnaires were analyzed. In the online group, 7 people started the questionnaire without completing it; there was no evidence of differential dropout by arm (<3 drop-outs per condition). In the clinics, 15 started the survey but did not complete it. In the clinics, the respondents started the study with the understanding that they would quit if called into their doctor, and all of the dropouts were in fact caused by being called into an appointment. There was no differential dropout by arm (2 - 5 per arm).

The two samples were similar in age and sex distributions (Table 7.2). Virtually all scored well on the S-TOFHLA health literacy measure, probably because participants had to have sufficient computer skills to complete a computer-based questionnaire. Thus, health literacy was not used as a variable in any analysis. However, the clinic sample had a much lower average education level, reported less computer use, and were more likely to score poorly in numeracy. Most (66%) were Hispanic. Clinic participants had slightly worse current health status, but were not more likely to report histories of flu, heart disease, or drug side-effects (three health issues related to the questionnaire scenarios). Clinic participants took several minutes longer on average to complete the survey, which could be attributable to their lower computer familiarity and numeracy, or to distractions in the busy waiting areas. In the clinics, those who participated were more likely to be young and English-speaking than the clinic populations as a whole. Several older respondents asked their children to translate individual words for them, and in one case to help input answers into the computer. One respondent with a bandaged arm had a friend input
There were no substantive demographic differences between the four experimental groups (random, sequential, switch, and search; data not shown).

### Table 7.2: Characteristics of on-line and clinic study samples

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>On-line (n=100)</th>
<th>Clinic (n=65)</th>
<th>p</th>
<th>Total sample (n=165)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age, yrs (range)</td>
<td>32.8 (19-61)</td>
<td>30.7 (18-72)</td>
<td>.90</td>
<td>32.0 (18-72)</td>
</tr>
<tr>
<td>Number (%), women</td>
<td>64 (64.0)</td>
<td>41 (63.1)</td>
<td>&gt;.99</td>
<td>105 (63.6)</td>
</tr>
<tr>
<td>Educational level, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no bachelor's degree*</td>
<td>19 (19.0)</td>
<td>28 (45.0)</td>
<td>&lt;.001</td>
<td>47 (28.5)</td>
</tr>
<tr>
<td>some college</td>
<td>37 (37.0)</td>
<td>23 (35.4)</td>
<td>.61</td>
<td>60 (36.4)</td>
</tr>
<tr>
<td>bachelor's or graduate degree</td>
<td>44 (44.0)</td>
<td>14 (21.5)</td>
<td>.13</td>
<td>58 (35.2)</td>
</tr>
<tr>
<td>Self-identity, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African-American</td>
<td>10 (10.0)</td>
<td>10 (15.4)</td>
<td>&lt;.001</td>
<td>20 (12.1)</td>
</tr>
<tr>
<td>Asian</td>
<td>20 (20.0)</td>
<td>0</td>
<td>.02</td>
<td>20 (12.1)</td>
</tr>
<tr>
<td>white</td>
<td>60 (60.0)</td>
<td>6 (9.2)</td>
<td>.001</td>
<td>66 (40.0)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>2 (2.0)</td>
<td>43 (66.2)</td>
<td>&lt;.001</td>
<td>45 (27.3)</td>
</tr>
<tr>
<td>other</td>
<td>3 (3.0)</td>
<td>3 (4.5)</td>
<td>.64</td>
<td>6 (3.6)</td>
</tr>
<tr>
<td>mixed race/ethnicity</td>
<td>5 (5.0)</td>
<td>3 (4.5)</td>
<td>.81</td>
<td>8 (4.8)</td>
</tr>
<tr>
<td>Mean health status ± SD</td>
<td>4.0 ± 0.7</td>
<td>3.7 ± 0.8</td>
<td>.002</td>
<td>3.9 ± 0.7</td>
</tr>
<tr>
<td>(1 = poor, 5 = excellent)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reported health history, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>history of flu</td>
<td>73 (73.0)</td>
<td>42 (64.6)</td>
<td>.46</td>
<td>114 (69.1)</td>
</tr>
<tr>
<td>diagnosis of heart disease</td>
<td>5 (5.0)</td>
<td>2 (3.1)</td>
<td>.81</td>
<td>7 (4.2)</td>
</tr>
<tr>
<td>history of drug side effects</td>
<td>53 (53.0)</td>
<td>24 (36.9)</td>
<td>.13</td>
<td>77 (46.9)</td>
</tr>
<tr>
<td>Computer questions, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>use every day</td>
<td>98 (98.0)</td>
<td>37 (56.9)</td>
<td>&lt;.001</td>
<td>135 (81.8)</td>
</tr>
<tr>
<td>very comfortable using mouse</td>
<td>96 (96.0)</td>
<td>56 (86.2)</td>
<td>.04</td>
<td>152 (92.1)</td>
</tr>
<tr>
<td>have no email address</td>
<td>0</td>
<td>7 (10.8)</td>
<td>.001</td>
<td>7 (4.2)</td>
</tr>
<tr>
<td>Numeracy category, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>poor (&lt;5 out of 8)</td>
<td>16 (16.0)</td>
<td>32 (51.6)b</td>
<td>&lt;.001</td>
<td>48 (29.6)b</td>
</tr>
<tr>
<td>adequate (&gt;5 out of 8)</td>
<td>84 (84.0)</td>
<td>30 (48.4)</td>
<td>.12</td>
<td>114 (70.4)</td>
</tr>
<tr>
<td>STOFHLA category, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>adequate health literacy</td>
<td>99 (99.0)</td>
<td>56 (98.2)c</td>
<td>&gt;.99</td>
<td>155 (99)</td>
</tr>
<tr>
<td>marginal health literacy</td>
<td>1 (1.0)</td>
<td>1 (1.8)</td>
<td>.04</td>
<td>2 (1)</td>
</tr>
<tr>
<td>Mean minutes to complete ± SD</td>
<td>14.2 ± 7.2</td>
<td>17.2 ± 5.3</td>
<td>.004</td>
<td>15.4 ± 6.7</td>
</tr>
</tbody>
</table>

*Chi-squared tests used for categorical variable and t tests for continuous ones.

a. Includes less than high school, high school graduate, and technical school
b. 3 clinic respondents missing numeracy scores because of interruptions
c. 8 clinic respondents missing STOFHLA scores because of interruptions
7.3 Preliminary analyses of numeracy and risk perception responses

7.3.1 Modification in the assessment of numeracy

During analysis, I discovered that I had inadvertently omitted one question from the numeracy questionnaire (questions are listed in Table 7.3).

### Table 7.3: Percentages answering correctly in three studies

<table>
<thead>
<tr>
<th>Question</th>
<th>Lipkus et al (n=463)</th>
<th>Schwartz et al (n=287)</th>
<th>on-line (n=100)</th>
<th>clinic (n=62)*</th>
<th>total (n=162)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imagine that we flip a fair coin 1,000 times. What is your best guess about how many times the coin would come up heads?</td>
<td>Not scored</td>
<td>54.0</td>
<td>74.0</td>
<td>66.1</td>
<td>71.0</td>
</tr>
<tr>
<td>Which of the following numbers represents the biggest risk of getting a disease? 1 in 100, _1 in 1000, _1 in 10</td>
<td>78.2</td>
<td>81.0</td>
<td>54.8</td>
<td>71.0</td>
<td></td>
</tr>
<tr>
<td>Which of the following numbers represents the biggest risk of getting a disease? 1%, 10%, 5%</td>
<td>83.8</td>
<td>92.0</td>
<td>80.6</td>
<td>87.7</td>
<td></td>
</tr>
<tr>
<td>If Person A's risk of getting a disease is 1% in ten years, and person B's risk is double that of A's, what is B's risk?</td>
<td>90.5</td>
<td>96.0</td>
<td>71.0</td>
<td>86.4</td>
<td></td>
</tr>
<tr>
<td>If Person A's chance of getting a disease is 1 in 100 in ten years, and person B's risk is double that of A's, what is B's risk?</td>
<td>86.6</td>
<td>question omitted</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 100?</td>
<td>80.8</td>
<td>95.0</td>
<td>67.7</td>
<td>84.6</td>
<td></td>
</tr>
<tr>
<td>If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 1000?</td>
<td>77.5</td>
<td>89.0</td>
<td>61.3</td>
<td>78.4</td>
<td></td>
</tr>
<tr>
<td>If the chance of getting a disease is 20 out of 100, this would be the same as having a ___% chance of getting the disease.</td>
<td>70.4</td>
<td>94.0</td>
<td>53.4</td>
<td>78.4</td>
<td></td>
</tr>
<tr>
<td>The chance of getting a viral infection is 0.0005. Out of 10,000 people, about how many of them are expected to get infected?</td>
<td>48.6</td>
<td>60.0</td>
<td>32.3</td>
<td>49.4</td>
<td></td>
</tr>
</tbody>
</table>

*Four of the 62 clinic respondents did not have numeracy scores because of interruptions*

In the Lipkus study (19), the first question in Table 7.3 was an unscored practice question; however, Schwartz et al (18) and Gurmankin et al (20, 222) have used the same item as a valid numeracy question. On an exploratory basis, I scored this "coin flip" question to bring the total number of items in my scale to 8 and found the resulting response distribution very similar to the one reported by Lipkus et al (Table 7.4, chi-square test for homogeneity, p=.26). In particular, the proportion of respondents with 5 or fewer correct was 29.6%, virtually identical to the 31.9% reported by Lipkus. Internal reliability of this modified scale (Cronbach's alpha = 0.70) was equivalent to that reported in the original (0.70 to 0.75 in different samples).
Table 7.4: Distributions of correct responses in two studies

<table>
<thead>
<tr>
<th>N of correct responses</th>
<th>Lipkus et al Percent</th>
<th>Ancker Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.9</td>
<td>0.6</td>
</tr>
<tr>
<td>1</td>
<td>2.0</td>
<td>2.5</td>
</tr>
<tr>
<td>2</td>
<td>4.0</td>
<td>4.3</td>
</tr>
<tr>
<td>3</td>
<td>5.0</td>
<td>4.3</td>
</tr>
<tr>
<td>4</td>
<td>8.0</td>
<td>4.3</td>
</tr>
<tr>
<td>5</td>
<td>12.0</td>
<td>13.6</td>
</tr>
<tr>
<td>6</td>
<td>13.0</td>
<td>19.8</td>
</tr>
<tr>
<td>7</td>
<td>24.0</td>
<td>25.9</td>
</tr>
<tr>
<td>8</td>
<td>32.0</td>
<td>24.7</td>
</tr>
</tbody>
</table>

Although this modified scale cannot be considered exactly equivalent to the original Lipkus scale, it appears very similar to the original and to modifications published by others (20, 222) and was adopted for all analyses in this study. For categorical analyses, respondents scoring 5 or fewer correct out of 8 (i.e., 62.5% correct) were considered to have low numeracy; nearly 1/3 of participants were identified as low numeracy by this criterion (Table 7.2).

Numeracy was modestly correlated with S-TOHFLA score ($r = 0.28$, $p < .001$), education ($r = 0.31$, $p < .001$), and health status ($r = 0.19$, $p < .001$). However, low numeracy remained prevalent in the highest health literacy group (found among 28% of those with adequate health literacy) and educational groups (16% of college-educated respondents, and 26% of those with some college).

### 7.3.2 Factor analysis of perceived risk questions

Factor analysis was performed to determine whether the four risk perception questions were measuring a single underlying construct. The susceptible and vulnerable questions were reverse-coded so that all questions would have low values indicating low perceived risk, and high values indicating high perceived risk. After principal components extraction without rotation, a two-factor solution accounted for 80.4% of the variance in Story 1 and 84.8% of the variance in Story 2.

Table 7.5: Factor loadings, Story 1

<table>
<thead>
<tr>
<th>Story 1 question</th>
<th>Component 1</th>
<th>Component 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Susceptible (reverse-coded)</td>
<td>.802</td>
<td>-.392</td>
</tr>
<tr>
<td>Vulnerable (reverse-coded)</td>
<td>.758</td>
<td>-.505</td>
</tr>
<tr>
<td>Verbal scale</td>
<td>.749</td>
<td>.493</td>
</tr>
<tr>
<td>Numeric estimate</td>
<td>.779</td>
<td>.420</td>
</tr>
</tbody>
</table>

Table 7.6: Factor loadings, Story 2

<table>
<thead>
<tr>
<th>Story 2 question</th>
<th>Component 1</th>
<th>Component 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Susceptible (reverse-coded)</td>
<td>.872</td>
<td>-.406</td>
</tr>
<tr>
<td>Vulnerable (reverse-coded)</td>
<td>.872</td>
<td>-.399</td>
</tr>
<tr>
<td>Verbal scale</td>
<td>.758</td>
<td>.398</td>
</tr>
<tr>
<td>Numeric estimate</td>
<td>.694</td>
<td>.577</td>
</tr>
</tbody>
</table>
All questions loaded heavily on the first factor, which may represent the general construct of perceived risk (perceived risk factor). The second factor revealed a sharp contrast between the first two questions and the last two, and may represent the distinction between feelings about risk and estimates of risk (feelings/estimates factor). A weak third factor (data not shown) contrasted between the verbal scale and the numeric estimate, with virtually zero loadings on the other two questions.

On the basis of this analysis, I summed the reverse-coded "susceptible" and "vulnerable" questions and subtracted 1 to create a combined risk feelings question for each story, with a score of 1 indicating low feelings of risk, and 7 indicating high feelings of risk. The verbal scale and the numeric scale, although apparently tapping into the same construct, were not combined, as this study focused heavily on numeracy and thus it was of interest to examine similarities and differences between verbal and numeric expressions. Also, the correlations between these two items were only 0.50 or less (see Section 6.4 below). I thus retained these as two separate items, verbal risk estimates and numeric risk estimates.

7.4 Primary outcomes: numeracy, graphics, risk perception, and intention

7.4.1 Numeracy and risk perception

Hypothesis A1. Low numeracy will be associated with higher risk perceptions and less accurate estimates with all graphics.

Among respondents without a college degree, lower numeracy score was correlated with higher risk feelings and numeric risk estimates for both scenarios (r's from -0.20 to -0.40, p's ≤ .04). Low education level (less than a bachelor's degree) was also associated with higher risk feelings and numeric risk estimates, and the correlations between numeracy and risk perceptions were much weaker or non-existent among those with bachelor's or advanced degrees.

Low-numeracy respondents were less likely than high-numeracy ones to choose the scenario-provided risk level for their numeric risk estimate; 67% of low-numeracy respondents
gave a non-scenario risk estimate for both stories, compared to 47% of high-numeracy ones (p=.02). Again, the effect was weaker among college-educated respondents.

For the quantitative risk estimate, there was a bias in favor of 50% (11% of respondents in Story 1, and 5% of respondents in Story 2 chose 50%). Low-numeracy respondents, even college-educated ones, were much more likely to use 50% (23% vs. 6% in Story 1; p=.002, 15% vs. 1% in Story 2; p <.001). This led significant differences in both means and variances of their risk estimates (Story 1: means of 42 vs. 30; p < .001 for significance of difference between means of square-root-transformed data, p<.001 for significance of differences between variances; Story 2: 22 vs. 6, p = .04 on square-root-transformed data; p<.001 for significance of differences between variances).

In Story 1, 41% of respondents gave a numeric risk of 29; the rest were fairly evenly split between those who said it was lower (26%) and those who said it was higher (33%). There was an interaction with numeracy: high-numeracy respondents were as likely to overestimate as underestimate, whereas low-numeracy ones were more likely than high-numeracy ones to overestimate (50% vs. 26%, p=.01).

In Story 2, 31% of respondents gave a numeric risk of 6; the rest were more likely to say it was lower (43%) than higher (26%). The interaction with numeracy was opposite to the one above: low-numeracy respondents were evenly split between overestimators and underestimators, while high-numeracy respondents were more likely than low-numeracy ones to lowball the risk (46% underestimates compared to 33%; p=.05).

For both stories, perceived risk ratings tended to be substantially higher among respondents with low computer familiarity, Hispanics, clinic respondents (vs. the online sample), and those with poor self-reported health status. For example, people with good or excellent self-reported health status were more likely to call the 6% risk “almost zero” or “very small” (67.2% vs. 37.8%, p = .007). Worse self-reported health was also associated with concern about side-effects (65.6% vs. 51.7%, p = .04).
7.4.2 Effects of graphics on risk perception

Hypothesis B1. Random graphics will be associated with higher perceived risk than sequential ones.

Hypothesis B2. Random graphics will be associated with more variability (lower accuracy) in risk estimates than sequential ones.

Hypothesis B3. The interactive switch graphic will produce perceived risk answers midway between the random and sequential ones.

Hypothesis B4. The interactive search graphics will be associated with higher feelings of risk for large risks, and smaller feelings of risk for small risks, than static graphics.

Group assignment (i.e., graphics) did not affect median risk feelings, verbal risk estimates, or numeric risk estimates, but it did affect variability in numeric estimates. For Story 2, variance in numeric risk estimates was highest for the random graphic \( s^2 = 630.2 \), lowest for the sequential graphic \( 105.9 \), and in the middle for the switch graphic \( 409.9 \) and for the search graphic \( 243.2 \). The variances were significantly different (Levene's test, \( F=4.7 \), \( p=.004 \)). The differences were attributable to more extreme outliers in the random group. In Story 1, the variances followed the same pattern but differences were smaller and not statistically significant.

However, an additional study of unlabeled graphics (Section 7.5.3 below) shows support for several of these hypotheses.

Hypothesis D1. There will be an interaction between the interactive search graphic and numeracy such that the interactive graphic will be associated with reduced differences in perceived risks between high- and low-numeracy participants.

For all groups, low-numeracy respondents reported higher perceived risks and stronger intention to get preventive action than high-numeracy ones did, especially for Story 2 (the 6% risk of heart disease). Differences between high- and low-numeracy respondents were large in the random and the switch groups, medium to small in the sequential group, and smallest in Group 4 (search graphics; Table 7.7).
Table 7.7: Risk perceptions and intention by numeracy in all groups and within Group 4 (search graphics)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean response</th>
<th>Adequate numeracy</th>
<th>Low numeracy</th>
<th>Difference in means (95% CI)</th>
<th>Low numeracy significantly higher?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All groups</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Story 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>risk feelings</td>
<td>4.25</td>
<td>4.77</td>
<td>0.52 (-0.07, 1.11)</td>
<td>borderline</td>
<td></td>
</tr>
<tr>
<td>verbal risk estimate</td>
<td>3.54</td>
<td>3.67</td>
<td>0.13 (-0.34, 0.60)</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td>numeric risk estimate</td>
<td>29.8</td>
<td>43.2</td>
<td>13.5 (4.8, 22.1)</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td>intention</td>
<td>2.34</td>
<td>2.06</td>
<td>0.28 (-0.06, 0.62)*</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td><strong>Story 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>risk feelings</td>
<td>2.54</td>
<td>3.54</td>
<td>1.01 (0.38, 1.64)</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td>verbal risk estimate</td>
<td>2.33</td>
<td>2.94</td>
<td>0.61 (0.19, 1.03)</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td>numeric risk estimate</td>
<td>6.3</td>
<td>22.6</td>
<td>16.4 (9.2, 23.6)</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td>intention</td>
<td>2.68</td>
<td>2.23</td>
<td>0.45 (0.10, 0.80)*</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td><strong>Group 4 only</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Story 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>risk feelings</td>
<td>4.45</td>
<td>4.60</td>
<td>0.15 (-1.11, 1.41)</td>
<td>borderline</td>
<td></td>
</tr>
<tr>
<td>verbal risk estimate</td>
<td>3.55</td>
<td>3.89*</td>
<td>0.34 (-1.24, 1.92)</td>
<td>borderline</td>
<td></td>
</tr>
<tr>
<td>numeric risk estimate</td>
<td>30.5</td>
<td>39.9</td>
<td>9.4 (-15.4, 34.1)</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td>intention</td>
<td>2.06</td>
<td>1.80</td>
<td>0.26 (-0.54, 1.06)*</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td><strong>Story 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>risk feelings</td>
<td>2.85</td>
<td>3.20</td>
<td>0.25 (-1.03, 1.73)</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td>verbal risk estimate</td>
<td>2.39</td>
<td>2.40</td>
<td>0.01 (-0.79, 0.81)</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td>numeric risk estimate</td>
<td>9.4</td>
<td>14.6</td>
<td>5.2 (-10.6, 21.1)</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td>intention</td>
<td>2.42</td>
<td>2.30</td>
<td>0.08 (-0.64, 0.82)*</td>
<td>y</td>
<td></td>
</tr>
</tbody>
</table>

*Difference in mean intention is calculated as adequate group – low group because on this 4-point scale, lower numbers indicate higher intention to take preventive action. For the other measures, lower numbers indicate lower risk and differences are calculated as low numeracy – adequate numeracy.

* n=9

7.4.3 Intention to take preventive action

**Hypothesis C1.** Random graphics will be associated with stronger intention to take preventive action.

**Hypothesis C2.** The interactive switch graphic will be associated with intention answers midway between those associated with the random and sequential ones.

**Hypothesis C3.** The interactive search graphics will be associated with stronger intentions to take preventive action for large risks, and weaker ones for small risks, than static graphics are.

Overall, the preventive medication was chosen by 62% of respondents in Story 1 and 49% of respondents in Story 2. Group assignment (graphics) had no appreciable effects on intention to take the preventive action.

In both stories, all answers about risk feelings, verbal and numeric risk estimates, self efficacy, response efficacy, and response side effects were correlated with intention to opt for the...
preventive action, in the expected directions, as predicted by the extended parallel process model. However, perceived disease severity was only weakly and non-significantly correlated with intention (Story 1: rho = 0.10; p=.20; Story 2: rho = 0.12, p = .14). In Story 1, the "realistic" and "accurate" questions were weakly but significantly correlated with intention (rho's <0.27) but not with risk feelings or risk estimates. The credibility measures (realistic and accurate, rho's < 0.40) correlated with self-efficacy. Self-efficacy also correlated with response efficacy (rho's ≥ 0.27, p's < .001).

Blacks and Hispanics were much more likely to opt for the preventive action than were whites or Asians (Story 1: 75%, 78%, 53%, 35%; p = .02; Story 2: 55%, 64%, 36%, 35%; p = .058). Also, the clinic respondents were more likely than the online sample to opt for the preventive action (71% vs. 56% for Story 1; p=.056; 60% vs. 41% for Story 2; p=.02). Racial differences partly mediated but did not fully account for the clinic effect (i.e., adding clinic to a logistic regression equation reduced the coefficient of the race variable and increased its p value). Low-numeracy respondents were somewhat more likely than high-numeracy ones to opt for the preventive action in Story 2 (60% vs. 44% for Story 2; p = .07) though not in Story 1.

Self-reported history of flu was not associated with likelihood of choosing the vaccine in Story 1, and self-reported history of heart disease was not associated with likelihood of choosing the preventive medicine in Story 2. Self-reported history of drug side-effects was associated with slightly lower likelihood of choosing the vaccine in Story 1 (53% vs. 68%, p = .07) but not in Story 2. Health status and educational level were not associated with likelihood of choosing the preventive action. There was a very weak association with age in Story 1 (the flu-like disease), with older respondents being slightly less likely to opt for the preventive action (correlation between age and intention: rho = 0.16, p = .04). For Story 1, women were somewhat more likely than men to choose the vaccine (67% vs. 53%, p=.09).

7.4.4 Effects of question order

For Story 2, in all groups, there was a strong question order effect. In the forward order (Figure 7.1), the questions about perceived risk appeared immediately after the interaction with the
graphic, and before the question about intention to take the preventive action. In the reverse order, questions about disease severity and efficacy appeared first before the intention question, with the perceived risk questions afterwards. Reverse-ordered groups were much more likely to opt for the preventive action (65% or more versus 34%; p < .001) in Story 2. There was no similar order effect for Story 1.

Question order had no appreciable effect on the perceived risk questions. However, it did have strong and statistically significant effects on self-efficacy, which was much higher in reverse-ordered groups (e.g., Story 1: 58.2% strongly agreed they could get the medication if they wanted it, compared to 41.9% p = 0.075). In Story 2 only, question order also had modest effects on response efficacy (higher in reverse-ordered groups) and concern about side effects (lower in reverse-ordered groups).

Figure 7.1: Order of question blocks

<table>
<thead>
<tr>
<th>Forward order</th>
<th>Reverse order</th>
</tr>
</thead>
<tbody>
<tr>
<td>View/interact with graphic</td>
<td>View/interact with graphic</td>
</tr>
<tr>
<td>A. Perceived risk</td>
<td>E. Severity</td>
</tr>
<tr>
<td>1. susceptible</td>
<td>11. disease severity</td>
</tr>
<tr>
<td>2. vulnerable</td>
<td></td>
</tr>
<tr>
<td>3. verbal risk</td>
<td></td>
</tr>
<tr>
<td>4. numeric risk</td>
<td></td>
</tr>
<tr>
<td>B. Attitudes toward graphics</td>
<td>D. Efficacy and barriers</td>
</tr>
<tr>
<td>5. realistic</td>
<td>8. self-efficacy</td>
</tr>
<tr>
<td>6. accurate</td>
<td>9. side effects</td>
</tr>
<tr>
<td>C. Intention</td>
<td>10. response efficacy</td>
</tr>
<tr>
<td>7. take preventive action?</td>
<td></td>
</tr>
<tr>
<td>D. Efficacy and barriers</td>
<td>B. Attitudes toward graphics</td>
</tr>
<tr>
<td>8. self-efficacy</td>
<td>5. realistic</td>
</tr>
<tr>
<td>9. side effects</td>
<td>6. accurate</td>
</tr>
<tr>
<td>10. response efficacy</td>
<td></td>
</tr>
<tr>
<td>E. Severity</td>
<td></td>
</tr>
<tr>
<td>11. disease severity</td>
<td></td>
</tr>
</tbody>
</table>

7.4.5 Effects of question order within search graphic group

_Hypothesis B4._ The interactive search graphics will be associated with higher feelings of risk for large risks, and smaller feelings of risk for small risks, than static graphics.

_Hypothesis C3._ The interactive search graphics will be associated with stronger intentions to take preventive action for large risks, and weaker ones for small risks, than static graphics are.
The order effect played out in an interesting fashion within Group 4 (the search graphic). For Story 2 with the 6% risk, the search graphics participants clicked from 2 to 51 times before finding a blue stick figure (median: 12 times). In the forward question order, there was an unusual effect: the more clicks before finding a blue person, the higher the perceived risk (correlation with verbal risk estimate, rho = 0.57; p = .009; risk feeling, rho = 0.35; p = .13). For example, those who clicked fewer than 10 times generally called the risk almost zero or very small, while almost all of those who clicked more than 25 times called the risk small to moderate. Increased risk perception correlated with increased intention to take protective action (verbal estimate rho = -0.44, p = .055; risk feeling rho = -0.74, p < .001; numeric estimate rho = -0.45, p = .047). However, as some participants had higher perceived risk than those in other groups and others had lower, the overall proportion in this group who opted for the protective action (30%) was no different from the proportion in other forward-ordered groups.

By contrast, in the reverse order, perceived risk was not correlated with click number (all rhos < .12). Risk perception remained correlated with intention, though more weakly than in the previous group, and, as noted above, 70% of this group opted for the preventive action.

In Story 1 (with the 29% risk), participants in the search group clicked a median of only 2 times before finding a blue figure (range: 1 to 16), and there were no correlations between click number and the perceived risk questions (all rho ≤ 0.07).

### 7.4.6 Modeling risk perception

These univariate analyses suggested a number of factors to be considered in any possible model of the perceived risk measures (risk feelings, verbal risk estimate, and numeric risk estimate). General linear models were constructed with each of the perceived risk questions in Story 1 and Story 2 treated as repeated measures.

**All respondents:** In backward and forward regression, group and question order were not significant contributors to any of the perceived risk questions. Although education, numeracy score, clinic status (vs. online), health status, and age were predictors of risk feeling, verbal risk estimate, and numeric risk estimate, no model using these predictors had an adjusted $R^2$ of larger
than 0.17, showing that these factors accounted for little of the observed variability in perceived risk.

Since the univariate analyses above suggested that the models would be different in the search groups, subset analyses were run on these groups.

- In the forward order, risk feeling was modeled well (adjusted $R^2$ for Stories 1 and 2 respectively were 0.50 and 0.77) by education (betas = −0.72 and −0.53), low computer familiarity (betas of 1.4 and 3.0), numeracy score (0.86 and 0.72), click number in Story 1 (−0.12, close to significant for risk feeling in Story 1) and number of clicks in Story 2 (0.04, significant for risk feeling in Story 2). Numeracy and low computer familiarity helped mediate the education effect, but did not fully account for it.
- In the forward order, verbal risk estimate could not be modeled well for Story 1 (adjusted $R^2 < 0.13$). However, for Story 2, a good model (adjusted $R^2$ of 0.57) was constructed from health status (beta = −0.67 for each increment in health status) and click number for Story 2 (beta = 0.04 per click).
- In the forward order, numeric risk estimate was modeled (adjusted $R^2$ of 0.30 and 0.26 for Story 1 and Story 2 respectively) by health status (−12.5 and −9.6), age (−1.0 and −0.40), and click number for Story 2 only (beta = 0.38 for Story 2).
- For the reverse order, a few factors (in-person vs. clinic, education, and numeracy) were statistically significant, they together accounted for less than 18% of the variation and thus were not considered strong predictors.

7.4.7 Modeling intention

General linear models were also constructed with intent as the outcome variable. For all groups, for Story 2, stepwise regression selected risk feelings (beta = −0.28), side effects (beta = −0.27), question order (beta = −0.45), and a small contribution from self-efficacy (beta = 0.17); adjusted $R^2 = 0.48$; $p < .001$. This was slightly better than a model in which group assignment was substituted for order. Substituting numeracy score for risk feelings also produced a good model
A generalized linear model suggested that there was a statistically significant interaction between concern about side effects and risk feeling, but this effect was very small. For Story 1, risk feelings (beta = -0.22), side effects (beta = -0.40), and rating on the accurate question (beta = 0.23) were the most important predictors (adjusted $R^2 = 0.42$).

Subset analysis was performed on the Group 4 (search) respondents (stepwise selection).

- For Story 1, in the forward order, risk feeling (beta = -0.46), severe (beta = -0.56), side effects (beta = 0.37), and race (beta = 0.21) were the strongest predictors of intent (adjusted $R^2 = 0.74$).
- For Story 1, in the reverse order, intent was predicted by risk feelings (beta = -0.34), side effects (beta = -0.46), and rating on the accurate question (beta = 0.42) (adjusted $R^2 = 0.79$).
- For Story 2, within the forward order, intent was predicted by risk feelings (beta = -0.30) and accuracy (beta = -0.36); adjusted $R^2 = 0.44$, $p = .003$.
- For Story 2, within the backward order, concern about side effects (beta = -0.46) was the only predictor of intent; adjusted $R^2 = 0.26$; $p = .009$.

### 7.5 Secondary outcomes

#### 7.5.1 Attitudes toward the graphics

In general, participants appeared to like the interactive graphics, particularly the searching graphics (Table 7.8). Although this version also earned more “confusing” ratings than the others, only 9% of users (all high-numeracy) considered them confusing.

People with poorer numeracy tended to consider graphics more helpful for understanding (i.e., in all groups, 53% of low-numeracy respondents strongly agreed graphics were helpful, compared to 42% of others; numeracy score correlated with better helpfulness rating, rho = 0.24, $p = .002$). People with lower numeracy were not more likely to consider graphics confusing (23% vs. 19%, $p = .56$). As a validity check, ratings on the helpful and confusing questions were negatively correlated (rho = -0.37, $p < .001$).
Table 7.8: Attitudes toward the graphics (univariate analyses)

<table>
<thead>
<tr>
<th>Strongly agreed that the graphic...</th>
<th>Group 1 (random)</th>
<th>Group 2 (sequential)</th>
<th>Group 3 (switch)</th>
<th>Group 4 (search)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 39)</td>
<td>(n = 44)</td>
<td>(n = 39)</td>
<td>(n = 43)</td>
<td></td>
</tr>
<tr>
<td>Helped me understand the risks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>all participants</td>
<td>13/39 (33%)</td>
<td>18/44 (41%)</td>
<td>17/39 (44%)</td>
<td>26/43 (60%)</td>
<td>.08</td>
</tr>
<tr>
<td>low numeracy only</td>
<td>6/13 (46%)</td>
<td>6/12 (50%)</td>
<td>6/13 (46%)</td>
<td>7/10 (70%)</td>
<td>.47</td>
</tr>
<tr>
<td>Was confusing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>all participants</td>
<td>1/39 (3%)</td>
<td>0</td>
<td>1/39 (3%)</td>
<td>4/43 (9%)</td>
<td>.13</td>
</tr>
<tr>
<td>low numeracy only</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1/10 (10%)</td>
<td>.10</td>
</tr>
<tr>
<td><strong>Graphic in Story 1 (29% risk)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Was realistic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>all participants</td>
<td>6/39 (16%)</td>
<td>14/44 (32%)</td>
<td>13/39 (33%)</td>
<td>23/43 (53%)</td>
<td>.004</td>
</tr>
<tr>
<td>low numeracy only</td>
<td>4/13 (31%)</td>
<td>4/12 (33%)</td>
<td>3/13 (23%)</td>
<td>8/10 (80%)</td>
<td>.03</td>
</tr>
<tr>
<td>Was an accurate way of showing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>the risks</td>
<td>7/39 (18%)</td>
<td>15/44 (34%)</td>
<td>19/39 (49%)</td>
<td>18/43 (42%)</td>
<td>.03</td>
</tr>
<tr>
<td>all participants</td>
<td>4/13 (33%)</td>
<td>4/12 (33%)</td>
<td>4/13 (31%)</td>
<td>6/10 (60%)</td>
<td>.43</td>
</tr>
<tr>
<td>low numeracy only</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Graphic in Story 2 (6% risk)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Was realistic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>all participants</td>
<td>13/39 (33%)</td>
<td>18/44 (41%)</td>
<td>13/39 (33%)</td>
<td>21/43 (49%)</td>
<td>.42</td>
</tr>
<tr>
<td>low numeracy only</td>
<td>4/13 (31%)</td>
<td>6/12 (50%)</td>
<td>4/13 (31%)</td>
<td>6/10 (60%)</td>
<td>.39</td>
</tr>
<tr>
<td>Was an accurate way of showing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>the risks</td>
<td>12/39 (31%)</td>
<td>20/44 (46%)</td>
<td>18/39 (46%)</td>
<td>24/43 (56%)</td>
<td>.15</td>
</tr>
<tr>
<td>all participants</td>
<td>5/13 (39%)</td>
<td>6/12 (50%)</td>
<td>5/13 (39%)</td>
<td>6/10 (60%)</td>
<td>.69</td>
</tr>
<tr>
<td>low numeracy only</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Chi-square or Fisher exact tests

The low computer familiarity group was more likely to consider graphics confusing (31.2% vs. 18.3%, p = .004). However, the proportions were almost exactly the same in every group, suggesting that the search and switch graphics were no more likely to be confusing than the static ones.

7.5.2 Correlations between verbal and numeric risk descriptions

Participants were asked to estimate their risk both on a 7-item verbal scale ranging from "almost zero" to "almost certain" (verbal risk estimate) similar to that used in other numeracy assessments (e.g., (165)), as well as on a numeric scale from 0% to 100% (numeric risk estimate). For Story 1 (the 29% risk), the median numeric risk estimate was 29 (IQR: 25-40, range: 0-100). The modal verbal risk estimate was 4 ("moderate"). For Story 2, the modal quantitative risk estimate was 6 (IQR: 2-8; range: 0-94), and the modal verbal risk was 2 ("very small").
Although verbal and numeric risk expressions were strongly correlated (rhos = 0.53 and 0.59 in Story 1 and Story 2, p's < .001), the scatterplots below (Fig. 7.2) show that there was still a very wide variation in interpretations. For example, "moderate" risk in Story 1 corresponded to numeric values ranging from 10% to 90%, and in Story 2 from 0% to 66%. Conversely, a risk of 29% in Story 1 was described as anything from "almost zero" to "very large," and a risk of 6% in Story 2 was described as anything from "almost zero" to "moderate." The correlations were not substantively different in the low- and high-numeracy groups.

Figure 7.2: Correlations between verbal risk estimates and numeric ones

7.5.3 Estimating proportions in unlabeled graphics

**Hypothesis B1. Random graphics will be associated with higher perceived risk than sequential ones.**

**Hypothesis B2. Random graphics will be associated with more variability (lower accuracy) in risk estimates than sequential ones**

In this substudy with unlabeled graphics, estimates of random arrangements were significantly more inaccurate than estimates of sequential ones for very high proportions (>= 60%) and low proportions (<30%), but not for medium ones.
Table 7.9: Average inaccuracy in estimates of unlabeled graphics

<table>
<thead>
<tr>
<th>Percent depicted and sample size</th>
<th>Random arrangement</th>
<th>Sequential arrangement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean estimate</td>
<td>Mean inaccuracy (95% CI)</td>
</tr>
<tr>
<td>6 (n = 165)</td>
<td>8.7</td>
<td>2.7 (1.4 to 4.1)</td>
</tr>
<tr>
<td>29 (n = 165)</td>
<td>33.7</td>
<td>4.7 (2.6 to 6.8)</td>
</tr>
<tr>
<td>40 (n = 43)</td>
<td>41.6</td>
<td>1.6 (-3.1 to 6.2)</td>
</tr>
<tr>
<td>50 (n = 39)</td>
<td>53.3</td>
<td>3.3 (-1.2 to 7.8)</td>
</tr>
<tr>
<td>60 (n = 44)</td>
<td>66.0</td>
<td>6.0 (0.5 to 11.5)</td>
</tr>
<tr>
<td>70 (n = 39)</td>
<td>81.0</td>
<td>11.0 (5.0 to 17.1)</td>
</tr>
</tbody>
</table>

*p significance of difference from 0

Mean estimates were fairly close to the true proportions (Table 7.9). Random arrangements elicited higher mean estimates than sequential ones for almost all the graphs, and the difference was statistically significant for the 6% and 29% graphs.

Very few participants gave the same estimate when viewing the same graphic in the different arrangements. The exception was 50%, which more than 40 percent of viewers recognized in both arrangements. However, respondents' estimates of the same quantity in different arrangements were correlated (all r's above 0.43).

Table 7.10: Differences between estimates of random and sequential graphics showing the same proportion

<table>
<thead>
<tr>
<th>Percent depicted and sample size</th>
<th>Mean paired difference (95% CI)</th>
<th>p**</th>
<th>N. giving same answer for both</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 (n = 165)</td>
<td>2.2 (1.3 to 3.2)</td>
<td>&lt;.001</td>
<td>33 (20%)</td>
</tr>
<tr>
<td>29 (n = 165)</td>
<td>6.0 (3.8 to 8.1)</td>
<td>&lt;.001</td>
<td>14 (8%)</td>
</tr>
<tr>
<td>40 (n = 43)</td>
<td>-1.6 (-3.8 to 7.1)</td>
<td>.55</td>
<td>2 (5%)</td>
</tr>
<tr>
<td>50 (n = 39)</td>
<td>2.7 (-1.1 to 6.7)</td>
<td>.16</td>
<td>16 (41%)</td>
</tr>
<tr>
<td>60 (n = 44)</td>
<td>3.4 (-2.3 to 9.2)</td>
<td>.24</td>
<td>2 (5%)</td>
</tr>
<tr>
<td>70 (n = 39)</td>
<td>4.5 (-2.8 to 11.8)</td>
<td>.22</td>
<td>2 (5%)</td>
</tr>
</tbody>
</table>

*34 (87%) chose 50 for this graph
**Paired t tests

Similar conclusions can be drawn from Table 7.11. For 6%, 29%, 50%, and 60%, the random arrangement was associated with overestimates (although the effect was not statistically significant for 60%). For 6% and 29%, the sequential arrangement was associated with underestimates. For 50%, the sequential arrangement resulted in accuracy rather than underestimates.

For 40% and 60%, arrangement did not have a strong effect. For 70%, both arrangements were overestimated. However, the random arrangement was especially likely to be overestimated and was overestimated by almost all viewers.
Hypothesis A1. Low numeracy will be associated with higher risk perceptions and less accurate estimates with all graphics.

Better numeracy was weakly correlated with decreasing inaccuracy (defined as estimate – true answer) for 29% \( (r = -0.26, p=.001 \text{ for random}; r = -0.16, p=.04 \text{ for sequential}) \) and 6% \( (r = -0.17, p = .03 \text{ for random only}) \). Low-numeracy respondents gave higher mean estimates for all graphics than high-numeracy ones, and the differences were statistically significant for 6% random \((7.7 \text{ vs. } 11.6, p=.01)\), 29% random \((38.7 \text{ vs. } 31.5, p = .002)\), and 29% sequential \((31.7 \text{ vs. } 25.8, p=.005)\). Low numeracy was not associated with tendency to answer 50% for any graphic.

Twenty percent \((n = 33)\) of respondents were as highly accurate (defined operationally as within ± 10 percentage points on all estimates, and between 0 and 12 on the 6% estimates). Correlations between perceived risk and click number were not different in high and low accuracy groups. Accuracy as a term in the regression models described above was not significant and did not change other terms.

### 7.5.4 Cognitive analysis

Two people, both with bachelor’s degrees and adequate numeracy, completed the questionnaire while providing verbal think-alouds. Both were given a practice exercise of multiplying two numbers to practice providing cognitive think-aloud protocols, and were instructed to verbalize thoughts, not to explain reasoning. One completed the questionnaire in Group 1 \((random \text{ graphic})\) and the other in Group 4 \((search \text{ graphic})\). These data were examined especially for indications of salience of the graphic in responses to the questions.
In responding to the 11 questions about each scenario, respondent 1 (random graphic) displayed a strong quantitative orientation by making 12 explicit references to the numbers provided in the scenarios. Most inferences were connected with numbers. In one case, she reframed the 29% chance of disease as equivalent to a 70% chance of not getting the disease. She also made 4 verbal references to the graph, most accompanied by a mouse gesture toward or over the graph.

Responder 2 (search graphic) focused a larger proportion of comments on verbal issues in the scenarios, such as the descriptions of the severity of the disease and the side effects and elaboration or interpretations in light of her own experience. She read the story more slowly, repeating key phrases, before starting the questions. While responding to the questions, she made 10 references to or elaborations upon the verbal information. For example, she distinguished between "the" risk described by the doctor in the scenario, and her ("my") personal risk, which she thought would be lower in light of her own health status. There were only 6 references to numbers, and 2 to the graphic. However, while she was clicking on squares in the search graphic, the pitch of her voice rose after several clicks, apparently expressing surprise that she hadn't found a blue person yet, and when she did find a blue person, she gave a brief chuckle or "hah" sound.

Verbal responses indicated that the constructs targeted by each question were somewhat intermingled. For example, Respondent 1 referred to both the severity and the likelihood of disease while answering the vulnerability and susceptibility questions, suggesting that these questions may have tapped into the combined construct of perceived threat (Figure 2.1) rather than simply perceived susceptibility. Similarly, in responding to the self-efficacy question, Respondent 2 referred not only to the scenario's description of the vaccine as free but also to a personal experience in which she had tried to get the flu immunization recommended by her doctor and found that the clinics had run out, suggesting that this question may have tapped into not only the psychological construct of self-efficacy but also perceived availability of resources or credibility of the doctor character in the scenario.
In accordance with standard verbal protocol analysis procedures (156), respondents were instructed to verbalize thoughts as they arose rather than explain them or elaborate on them. Concurrent verbalization has higher validity than retrospective verbalization, and Ericsson and Simon have found that it does not change thought sequences (156). However, in three instances, my respondents changed or corrected their answers while verbalizing. For example, one respondent chose "strongly agree" for the statement that if she got heart disease, she would be seriously ill, then before continuing to the next question, said "People can have heart disease for years, can't they?" and changed her answer to "mildly disagree."
CHAPTER 8. DISCUSSION AND CONCLUSIONS

This dissertation has presented three sequential studies that contribute to the risk communication and informatics literature. First was a systematic review of risk graphics, resulting in the development of an ontology of risk graphic features and types. Second, a qualitative study was conducted that drew from the ontology and resulted in the development of several novel graphics. Finally, a questionnaire study was conducted to assess the effect of these novel graphics on risk perception and decision-making. These three studies were conducted in the context of a model of health risk communication incorporating modifications that are particularly appropriate to informatics research. This adapted model predicts audience receptivity to a health precaution message on the basis of features of the message (such as efficacy and susceptibility), plus features of the numbers and graphics used to illustrate and describe the risk, and the audience's cognitive skills (such as numeracy). The questionnaire data was consistent with this model, showing that numeracy was a significant predictor of perceived risk, which in turn was correlated with intention.

8.1 Systematic review and ontology

The systematic review shows an extensive literature on risk graphics, somewhat limited by its focus on single-feature manipulations of graphics. The body of work is sufficient to construct an ontology of the features of graphics that affect cognition, perception, or behavior, as well as a categorization of four purposes of risk communication: information, education, decision support, and persuasion.

However, the review also reveals gaps and suggests areas for additional work. There is still room for single-feature studies on some topics. For example, the evidence supporting the use of stick figure graphics (also called icon graphics) for low-literacy audiences is strong. However, it was not clear which of the various arrangements is optimal: icons arranged as a group (sequential) or scattered throughout the graphic (random). The Cleveland and McGill mode literature suggests that random icons might be judged on the basis of the overall hue or saturation of the graphic, the least accurate of the available modes (79, 80). Alternately,
according to the Hollands summation model (109), the random icons might be visually "summed" or added up, which is also a relatively inaccurate task. Both models support the hypothesis that estimates of proportions would be more accurate with sequential arrangement and less accurate with random arrangement. However, neither helps to predict how such an effect would affect learning, persuasion, or decisions. My questionnaire study (discussed below) allowed me to test (and confirm) the hypothesis about accuracy of judgment with random vs. sequential arrangements, as well as to assess the effect on decisions.

The review also suggested that far more work is warranted into animation and interactivity, particularly because they are now so widely available in computer-mediated communication. My review of findings from psychologists and educators suggests that to be effective, animations and interactivity should be fully integrated. Animations can be too overwhelming if they do not allow viewers to start, stop, and "interrogate" them to learn more about the phenomenon they portray. Conversely, interactivity without animation (e.g., interactive exchanges of verbal and numeric information as in the colorectal risk estimator described earlier) fails to exploit the rich possibilities of virtual experience. Procedural skills learned through virtual experience are transferable to the real world (130), and in the lab setting, interactive animations focusing on chance and risk had measurable effects on monetary decision-making (31, 32). These results suggested that similar interactive experiences could be harnessed to give computer users an experience of the probability of disease in health-related risk communication. The qualitative study discussed below used formative development methods to develop interactive animations focusing on the probability of disease in health communication. These provided experiences such as collecting virtual samples of people or clicking on images of people to see if they are affected or not affected by the disease. The subsequent questionnaire study showed that the interactive animations affected risk perception and decisions.

Finally, the review suggested that an important under-researched topic is the interaction between graphic design and viewers' numeracy. The literature shows that numeracy strongly affects viewers' comprehension and use of numerical information such as numerical risks, as well as trust in and comfort with that information. Education level (and possibly by extension
numeracy) also appears to have effects on ability to interpret some graphics. However, graphics researchers have not systematically explored effects of numeracy on interpretation and use of different types of graphics. In the questionnaire study presented here, I assessed numeracy as a covariate and confirmed that numeracy interacted both risk perception and decisions on the basis of 4 different types of graphics.

8.2 Qualitative study to develop novel interactive graphics

The qualitative study allowed me to pilot different types of interactive risk communication graphics based on the icon graph, reject types (such as repeated sampling) that were unlikely to be acceptable, and explore participants’ interpretations of what they were seeing and experiencing. I designed the study to examine usability, user preferences, and user interpretations, and to incorporate these findings into software modifications. The designs seemed to be highly usable. However, one of my interactive graphics, the repeated sampling graphic, did not appear to be very helpful or clear. The concept of sampling, familiar to people with statistical training, might be too foreign to everyday lay conceptions of chance. More concrete visual metaphors might be helpful, such as bird’s eye views of cities or sports arenas. Alternately, it is possible that the graphics were successful in demonstrating sampling but unsuccessful in making this concept seem relevant to a consumer’s personal health risks and goals. By contrast, the game-like interaction of clicking on squares to look at the people underneath seemed to be more engaging and interesting. It is possible that this interaction encouraged users to think about how the abstract information applied to themselves.

This interactive software appeared to produce an affective response not found with other risk graphics. Others have argued that the feeling of risk, rather than solely its perceived quantitative magnitude, is an essential component of risk-related decision-making (42, 43). The feeling of risk, or perceived threat, is an essential component of health behavior models such as the health belief model (1), the precaution adoption process model (2), and the extended parallel process model (3, 10). A person’s feeling of risk predicts subsequent influenza vaccination better than that person’s perception of the risk magnitude (65). Methods of inducing positive affect have also
been explored in therapeutic contexts (223). Thus, a risk communication modality that carries an affective impact could potentially be useful in increasing perceived threat of high risks, calming fears about low risks such as adverse effects of vaccines, or helping patients compare risks. A potential application might be in patient-oriented decision support involving risk tradeoffs, such as the choice between cancer therapies. In these situations, there are no risk-free options, and the patient may be asked to make a choice based upon his or her feelings about the various risks. The findings may be particularly applicable in providing decision support for people with poor numeracy skills, in light of research showing that people with poor numeracy are less likely to derive meaningful affective information from numbers and numerical operations (55). It is, however, important to consider the potential for misusing affective risk communication. For example, in health promotion, increasing perceived threat can increase motivation to take protective action, but it can also inhibit action if respondents believe the action will not be effective against the threat (3, 10). Thus, in health promotion, efforts to increase perceived threat should be accompanied by efforts to increase the perceived efficacy of actions against the threat. Also, reducing perceived threat may not be ethical if it causes patients to ignore legitimate health risks, such as the risks of adverse effects of a medication.

As other studies have shown (23, 81, 86), participants suggested that the sequential arrangement of stick figures (Figure 1A) was better for expressing the magnitude of the risk. Cognitive psychology studies suggest that when an illustration shows a shape with a portion highlighted, judging proportions is an automatic and fairly accurate process (194). My finding that focus group participants judged the random arrangement (Figure 1A) more “realistic” is consistent with recent work by Schapira and colleagues showing that the random arrangement was perceived as more “true” (26). Future work might explore the effect on the credibility of the risk communication, which is relevant in light of findings that patients often reject the results of risk calculators (21). Schapira et al also found that the random arrangement increased the size of the estimated risk (26), but I did not assess quantitative judgments in this study. The participants’ preference for human figures as more personal than numbers or bar charts is similarly consistent with preferences found in other studies (26). The participants suggested that the random
arrangement was better for expressing unpredictability, or a sense that the disease might strike anybody, a finding that appears to be novel. Future work might examine whether perceived vulnerability is increased by a random arrangement.

Patients' well-documented difficulty with numerical information (18, 19, 22) might appear to suggest either that numerical risks should be avoided when possible in health communication, or that such numbers should always be accompanied by interpretations such as "high risk" or comparative information such as "higher than average." In fact, such qualitative descriptions are the only risk information provided by many health communication materials and studies (224); an example is the personal risk assessments on the Harvard School of Public Health's www.yourdiseaserisk.harvard.edu site. When people estimate their own personal risk of an illness, they tend to have a difficult time providing an absolute risk, and their relative or comparative estimates are better correlated with objective risk factors than their absolute estimates (21). Providing patients with absolute risks produced a better improvement in self-estimated absolute risk, whereas giving patients a combination of absolute and relative risks produced a better improvement in self-estimated relative risk (124). Worry about cancer is related independently to both absolute risk perception (the individual's estimated likelihood of getting cancer, measured on a qualitative scale from "very low" to "very high") and relative risk perception (how the individual believes his or her risk compares to that of others) (176). Gurmakkin et al found that consumers with high scores on a numeracy assessment tended to prefer and trust risk communications that included a numerical risk, although consumers with lower numeracy scores placed more trust in communications containing only verbal descriptions (20); a relationship between communication preference and computational accuracy was also found by Waters (5).

Nevertheless, when the focus group participants were using the interactive graphic, which showed absolute risk only, they did not ask for interpretation or comparative information. This might be an artifact of the rapid pace of the focus group discussions or other group effects, or it might indicate that such information was not needed with this risk communication modality. Future work might explore when qualitative or comparative descriptors are helpful in helping people derive meaning from absolute risk information or other types of quantitative estimates.
Distrust of medical research, as well as of the medical system, remains high among African-Americans; for many, the Tuskegee syphilis study has come to symbolize a hostile medical system (225, 226). Several of the focus group participants expressed distrust of doctors' motives or described medical drugs as risky. It is possible that such distrust might have contributed to the doctor-patient communication problems mentioned by the participants. The repeated references to information-seeking on the Internet underscore the need for on-line sources of reliable, trusted, and comprehensible information for the public, and particularly for the African-American community.

The study was limited by a small sample size and overrepresentation of women and of well-educated participants. This study also focused heavily on user preferences for communication format, which may promote patient engagement and satisfaction (212) but do not necessarily ensure comprehension (162), learning (102) or accurate decision-making (5, 30). The subsequent questionnaire study was designed to assess the effect of these graphics on such constructs as perceived risk, information credibility, and risk-related decision-making. In addition, this study qualitative study focused on comprehension of individual risks, not risk trade-offs or comparisons (5). The current version of the graphic has the limitation that it would be unlikely to be useful for expressing extremely rare risks, such as 1 in a million.

Nevertheless, this study provided guidance in the transition from laboratory finding to a useful consumer health communication application, and explored lay interpretations of interactive risk and probability graphics in health communication. It also suggested that interactive graphics may be a new medium for conveying comprehensible, credible, and motivational health and health promotion information, which may help improve health-related decision-making and outcomes.

8.3 Implications of questionnaire study

This questionnaire study focused on risk perceptions and intentions to take protective action when risks were illustrated with one of two interactive graphics (switch and search) or one of two static graphics (random and sequential). In the extended parallel process model, perceived risk is related to intention, but not in a simple way. Increasing perceived risk will lead to stronger
intention to take protective action only if perceived efficacy is high. If perceived efficacy is not high, increasing perceived risk may lead to rejection of the protective action. As a result, health promotion is sometimes best served by increasing perceived risk, and other times by reducing it or by increasing perceived efficacy.

The scenarios in the current study described protective measures that had very high response efficacy. Thus, the extended parallel process model predicted that increasing perceived risk would be correlated with increased intention to take the preventive action, a prediction that was supported by the data. In addition, the current study showed that numeracy strongly affected risk perceptions; that only the search graphic had an appreciable effect on risk perceptions; and that question order effects attenuated this finding. EPPM and other health behavior models (1, 10) suggest that people form risk perceptions when they process the severity and susceptibility information in the health message. However, the current findings demonstrate that numeracy affects perceived risk even when the severity and susceptibility information in the message are held constant. This supports the contention in Chapter 2 that theoretical models of health communication can be made more predictively powerful if they account for individual cognitive differences.

8.3.1 Numeracy

As I had hypothesized, numeracy had strong effects in this study. Although poorer numeracy was associated with lower educational level, the regression models suggest that the numeracy effects were not fully explained by education, age, computer literacy or other factors. Others have also found effects of numeracy independent of education, SAT scores, and other factors (18, 55).

In the scenario study, low-numeracy respondents systematically gave numeric risk estimates that were higher than the value described in the scenario, as well as higher than the estimates provided by other respondents. Low-numeracy respondents also reported higher risk feelings and higher verbal risk estimates than other respondents. Others have also linked low numeracy with overestimates of personal risk of disease (222, 227). In my study, the effect was in part because low-numeracy respondents were more likely to choose 50% when describing their own risk. This
may have been a rhetorical measure to express uncertainty or confusion (222, 228). However, this phenomenon did not fully account for the overestimation, because average estimates were above the scenario risk even among participants who did not choose 50%. Another reason low-numeracy respondents' estimates were higher than high-numeracy respondents' was that in the low-risk scenario, high-numeracy respondents tended to lowball the risk.

People with poor numeracy considered graphics more helpful for understanding the risk information, which is consistent with previous findings about 'subjective' numeracy (229, 230).

In the substudy of unlabeled graphics, better numeracy was correlated with better accuracy in estimating the proportion in the graphic, and low-numeracy respondents gave higher mean estimates for all graphics than high-numeracy ones. A number of other researchers have found that interpretation of risk graphics such as survival curves improves with either previous education or targeted training (25, 61). However, previous studies of simple tasks such as estimating proportions have been conducted in educationally homogeneous groups (college students (e.g. (109)) or have not assessed educational level or numeracy as a covariate (e.g., (162). It is possible that the computational skills measured by the Lipkus questions (19) contribute to the ability to interpret graphics. Alternately, the ability to estimate quantities visually could itself contribute to the development of the computational skills (17).

8.3.2 Effects of static graphics

I had hypothesized that among the static graphics, arrangement would affect risk estimates, with random arrangements appearing riskier than sequential ones and thus prompting more interest in adopting preventive behaviors. I also hypothesized that accuracy would be lowest with random graphics. These hypotheses were suggested by findings from my qualitative study (Chapter 6) which suggested that random arrangements might be perceived as more personally relevant (thus riskier) than sequential ones. Also, psychophysical findings discussed in Chapter 5 suggest that estimation tasks that require mentally summing non-contiguous blocks or lines (as in the random graphic) are less accurate than estimating proportions in lines or blocks as in the sequential one (80, 193, 194).
The hypotheses linking random graphics with higher perceived risk and stronger intention to take protective action were not supported by the scenario study, in which arrangement had no effect on mean numeric risk estimates, verbal risk estimates, risk feelings, or intentions. Nevertheless, in the substudy of unlabeled graphics, random arrangements elicited markedly and statistically significantly higher estimates than sequential ones, and the effect was largest for large proportions. Schapira has also recently found that random graphics elicited higher probability estimates (26). I conclude that a random arrangement does tend to be perceived as a larger proportion than a sequential one, but that on average this perception is not sufficient to inflate risk feelings or estimates when the graphic is used to illustrate a vivid verbal scenario.

As predicted, accuracy in estimation of the unlabeled graphics was lowest with the random arrangement, and even in the scenario study, variability in numeric risk estimate was significantly higher in the random arrangement because of an increased number of very high estimates. This lends support to the idea that the visual perception issue may have influenced the cognitive responses among some respondents. Additional evidence from the scenario study was that the interactive switch graphic, which allowed viewers to toggle back and forth between the random and sequential views of the same proportion, resulted in a numeric risk estimate variance that was midway between the variances associated with the random and sequential arrangements.

### 8.3.3 Question order effects

An unexpected finding was that question order had a strong effect on intention regardless of graphic. When questions about self-efficacy and response efficacy preceded the intention question (reverse-ordered questionnaire), most respondents said they would take protective action. But when questions about perceived risk preceded the intention question, most said they would not take the protective action (forward-ordered questionnaire). Witte proposes that consumers will tend to adopt a protective measure against a health threat (a) when the threat is large and relevant enough to evoke concern, but not so large as to evoke denial, and (b) when response efficacy and self-efficacy appear high enough to successfully address the threat (10, 41, 44). The reverse-ordered questionnaire may have increased intention to adopt the protective
behavior by heightening respondents' perceptions of response efficacy and self-efficacy, both of which were designed to be relatively high (the recommended response was described as very effective and free). In support of this explanation, self-efficacy was higher in the reverse-ordered groups.

An alternate explanation is that the forward ordering, by drawing attention first to perceived risk, may have elevated perceived risk to counterproductive levels, leading to rejection of the health promotion message, according to Witte's theory of failed fear appeals (10). However, this is a less plausible explanation, because even in forward-ordered groups and the higher risk story (Story 1), intention was strongly correlated with increasing risk perception. Scatterplots (not shown) showed no U-shaped relationships between the perceived risk questions and intentions, suggesting that perceived risk had not reached this counterproductive ceiling in this study.

8.3.4 Effect of interactive search graphic

I had hypothesized that the search graphic would focus attention on the relationship between the numerator and denominator in the risk, thus reducing risk perceptions for rare events and increasing risk perceptions for large ones. The idea of drawing attention to the denominator of a ratio has been suggested as a way to combat biases such overreaction to rare events (85) and ratio bias (184), the well-documented finding that people tend to perceive a chance of 10/1000 as larger than a chance of 1/100. In addition, the search graphic had some similarities with stimuli described in (32, 96), in which subjects learned about risks through the experience of sampling repeatedly rather than by interpreting symbolic (verbal, numeric, or graphic) descriptions. These experiments had found that experiential learning was associated with underestimation of rare events.

However, in my search graphic group, the more the respondent searched before finding a stick figure affected by the disease, the higher the perceived risk.

Sampling bias would be a plausible explanation for this phenomenon if I had not designed the game to force every respondent to find a blue figure. That is, if respondents had been able to stop clicking whenever they wanted, those who drew small samples would have been less likely to find
any blue figures (particularly in the 6% scenario); respondents who clicked only a few times would tend to have a reduced perception of the risk, and those who clicked many times would have a more accurate perception. (The Weber and Hertwig studies used this paradigm.\textsuperscript{1}) However, in my design, every sample of clicks necessarily contained exactly 1 blue figure, because the respondent had to continue clicking until a blue figure was discovered. Thus, a sample of 3 clicks contained exactly 1 blue figure (for an apparent probability of 33%), and a sample of 50 clicks contained exactly 1 blue figure (for an apparent probability of 2%). Nevertheless, perceived risks were reported as higher in the latter case.

Several explanations are possible.

(1) \textit{Vividness or availability}: Respondents were told that each square concealed a person who might or might not have the disease. If this instruction led the respondent to imagine the disease at each click, then a long sequence of clicks might lead to a more vivid or more available conception of the hazard, which could translate into a higher perceived risk. Others have found that describing risks in frequencies (e.g., 10 in 100) brings to mind more vivid imagery of the risk than describing risks as percentages (55, 64). Evidence that might support this explanation was that in Story 2, severity was also modestly correlated with click number, although the correlation was small and not statistically significant.

(2) \textit{Emotional arousal}: The qualitative study (Chapter 6) showed that the clicking interaction in the search graphic was associated with more verbal expressions of emotion than any static graphic, which suggests emotional arousal as a potential explanation. Also, the search graphic is similar to the Columbia Card Task, a game used in psychological experiments on risk-taking in which players click on virtual cards to see if each carries a monetary gain or a loss. In this game, clicking on the cards is associated with skin conductance responses, which indicate emotional arousal (231). A long sequence of clicks might have a cumulative effect on emotional arousal and feelings of risk. Indirect evidence consistent this explanation was the fact that the effect

\textsuperscript{1}Another difference between this study and experiments described in Weber et al 2004 and Hertwig et al 2004 was that our respondents sampled only once: that is, they clicked on squares until they found a blue person, then stopped. By contrast, the Weber and Hertwig studies describe subjects sampling repeatedly over time, and thus having a lengthier opportunity for associative learning.
disappeared in the group of respondents who first answered questions directing their attention to other (non-risk-related) dimensions of the problem; this question order may have postponed their consideration of the risk perception until after their heightened emotional arousal had worn off.

(3) **Gambler's fallacy:** In the gambler's fallacy (232, 233), a person who encounters a short-term deviation from the expected probability of random events becomes convinced that the probability will change in the future to maintain the expected average probability. According to this belief, the fact that 2 airplanes crashed in the U.S. in January reduces the chance of any crashes for the rest of the year, and a coin that came up tails 4 times in a row is more likely than not to come up heads the next flip. A long sequence of clicks on 'healthy' stick people could lead participants to believe that the probability of disease for future clicks must rise. This belief would inflate the perceived probability associated with (hypothetical) future clicks, which might also have carried over to the perceived risk questions.

In the case of truly independent random events such as coin flips, this belief is a fallacy because the events are independent and have no effect on each other. This belief would also have been a fallacy if each of the buttons in my graphic had a fixed and independent probability of disease, producing a binomial distribution of diseased stick figures. However, as described in the Methods chapter, I designed each graphic to contain a fixed number of diseased figures. Ironically, that means the gambler's fallacy is not entirely fallacious in this situation. As determined by the hypergeometric distribution (see Table 4.1 in Methods, truncated as Table 8.1 below), the chance of a diseased figure does in fact rise after a sequence of clicks on non-diseased stick figures. Nevertheless, as shown in Table 8.1, the probability rises extremely slowly, and even after 50 yellow clicks has risen from 6% to just under 7.5%, so the inflation in perceived risk was probably disproportionate to the actual increase in probability.
In this experiment, I did not rig the interactivity, so every respondent experienced a different number of clicks. However, the effect of clicking on perceived risk and intention suggests that a graphic similar to this one could be used to promote adoption of preventive measures for rare events if it were designed to ensure a large number of clicks.

### 8.3.5 Effect of search graphic on the less numerate

As described above, low-numeracy respondents typically reported much higher risk perceptions and stronger intention to take preventive action than high-numeracy respondents did. However, the interactive search graphics nearly eliminated these differences between high- and low-numeracy respondents by reducing perceived risk among the less numerate, and also slightly increasing perceived risk among the more numerate (Table 7.7). This effect suggests that graphics such as these could improve communication by reducing differences between the way that numerate health care professionals and less numerate patients perceive risks. It also suggests that graphics such as these might be particularly effective when the goal is to reduce overreactions to risks among the less numerate, such as situations in which high perceived risks lead people to reject health promotion messages.

In the current study, it is possible that the discrepancy between high- and low-numeracy respondents was attributable to the way they processed or attended to the different pieces of information in the message. Peters has found that the less numerate are not only less able to reason analytically about numbers, but also less adept at making affective judgments about them.
For example, when asked how “clear” a feeling they had about whether a particular option had a good or bad chance of winning, low-numeracy subjects reported less clear feelings about both the probability information and the pay-off amount. This suggests that in the current study, high-numeracy respondents may have relied on the percentages and the graphic to determine their feelings about perceived risk and response efficacy. By contrast, low-numeracy ones may not have gotten clear feeling from the number and instead may have relied on the descriptive text. The search graphic, however, forced all respondents—both high- and low-numeracy—to explore the probability through the interaction with the graphic. It is possible that this led both high- and low-numeracy groups to rely equally on the numeric information.

Another possible explanation is that with the interactive graphics, probabilities might have been learned through association (i.e., the association between clicking and the blue figures) rather than more effortful learned strategies (available primarily to the numerate), thus diminishing the differences between high- and low-numeracy respondents. This explanation has been proposed as an explanation for effects of a virtual card game, as well as for communication difficulties between doctors (who learn about rare events through experience) and patients (who learn about them primarily through description) (31). However, this explanation appears less likely in the experiment because of the negative correlation between click number and perceived risk; associative learning might be expected to produce a positive correlation.

8.3.6 Attitudes toward graphics

The search graphics were given higher ratings on helpfulness, accuracy, and realism than the other graphics, and ratings among low-numeracy respondents were particularly high (at least when sample sizes were sufficient to draw reliable conclusions: Table 7.8). Also, respondents with low computer familiarity were no more likely to consider the interactive graphics confusing than they were to consider static graphics confusing. The questions were not adopted from any previously validated scale, so it is possible that they measured something other than the intended construct. For example, they might indicate that the search graphics attracted more attention,
created more emotional engagement, or were more memorable than the others. In any case, the findings suggest that these interactive graphics were viewed positively by the respondents, and these graphics or similar ones may help to increase engagement with quantitative information in health contexts.

8.3.7 Strengths and limitations

A limitation of the questionnaire study was that it used text descriptions of hypothetical health choices, which were clearly less meaningful than real choices would be. Although the scenarios asked participants to imagine that the story had happened to them and that a physician had calculated a personalized risk for them, it is unlikely that participants reacted the way they would have for a genuinely personalized risk estimate. In addition, although intentions in general are linked to behavior (234), this experiment was not able to confirm whether intentions expressed here would predict future behavior.

Integrity of on-line questionnaire research and comparability between on-line and in-person samples are issues of concern in informatics (235), public health (e.g., (236), and psychology (237). Large on-line samples can be recruited quickly and cheaply to boost statistical power, but they also pose potential problems. External validity (generalizability) would be compromised if the sample reflects not the general population but the on-line one, which is somewhat more likely to be young, white, and well-educated than the average American (238). In fact, my on-line sample was relatively well educated (44% had a college or graduate degree), and its racial diversity largely resulted from an overrepresentation of Asian respondents (20%). Adding the clinic sample clearly improved external validity by broadening the range of education, numeracy, computer literacy, and health status levels represented. As a result, the participants were significantly more diverse than samples in other studies of health numeracy such as (20), who had a mean of 15 years of education, and (19), of whom only 12% had a high school education or less.

Nevertheless, internal validity could be threatened if responses from the on-line sample were systematically different from responses in the in-person sample because of issues such as different amounts of experimenter control over the environment or differential dropout rates that
could bias assignment. In fact, the clinic waiting rooms had many distractions, which may have caused the somewhat longer questionnaire duration among clinic participants. Also, dropout rates were much higher in the clinics than on-line (19% vs. 6%), but this was because clinic respondents started the study with the understanding that they would quit if called by the doctor for whom they had been waiting. In all cases, dropouts were caused by this event, and there is no evidence of differential dropout rates by experimental arm in either sample. Internal validity could also be compromised if the two samples demonstrated differences not attributable to measured factors such as education or health status. When I controlled for these factors, in-person status (vs. on-line status) almost entirely disappeared as a predictive factor. It did remain statistically significant in one of the analyses of intention, but it was clearly not clinically significant as the beta and the coefficient of determination for the entire equation were both extremely small (adjusted \( R^2 < 0.15 \)). I conclude that using two samples improved the representativeness of the samples without compromising internal validity.

A related generalizability concern was that participation in the study was restricted to English-speaking participants with a certain minimum level of computer literacy, so applying conclusions to other populations may be limited. All but one of my participants tested in the top level ("adequate") of health literacy, perhaps because the requirement to complete the questionnaire on a laptop excluded people with poor literacy. It is, however, encouraging that the small subgroup with less computer literacy did not give markedly different responses on any questions, and in particular did not rate the interactive graphics as more confusing than any other graphics.

The restricted sample size means several interesting subgroups had too few members for subset analyses. For example, I could not do an extensive analysis of the effect of different click numbers within the less numerate because only 9 of the search graphic group had low numeracy. Screening for numeracy and conducting additional studies within this sub-population would allow closer examination of ways to compensate for this factor. In addition, although the unlabeled graphics portion of this experiment explored a wide range of probabilities (from 6% to 80%), most of the focus was on the 6% and 29% scenarios. The experiments thus shed little light on the difficult issue of communicating about extremely small risks (85, 110) or very high ones.
The modifications to the S-TOFHLA and the numeracy scale should be kept in mind when comparing the results of this study with the results of other studies. The online S-TOFHLA created for this study has not been validated against the paper version. In addition, the modified numeracy measure differed from the original Lipkus version by 2 questions. However, as described above, the revised numeracy measure displays excellent scale reliability properties and a response distribution very similar to the original's. In addition, there is no consensus on the best way to measure numeracy and a number of alternative scales are available (18, 229, 230). Furthermore, others have modified the Lipkus instrument and other available numeracy instruments (e.g., (166)) without explanation or any demonstration of scale reliability (20, 222).

The think-aloud experiment was too small to be more than exploratory, but it does suggest that salience of the numbers and graphics varies from individual to individual, even within groups with adequate numeracy. Others have suggested that interest in quantitative medical data, at least partly independent of numerical ability, influences health decision making and information seeking (239). It also demonstrates that the questions do not succeed in targeting pure psychological constructs, but rather that people answer many questions in light of several related concepts. Chapter 3 identifies several mechanisms of representation effects, including manipulations of salience, cognitive load and knowledge demands, and learning effects. Analysis of larger numbers of respondents could potentially distinguish some of the mechanisms whereby the interactive search graphic affected risk perception, as well as the issues underlying the question order effect.
8.4 Conclusions and future work

As Figure 8.1 suggests, risk graphics are an important component of risk communication and are thus integral to both health communication and consumer informatics applications such as patient portals and consumer websites. The current findings have immediate applications in these domains. They support the use of sequential stick figure graphics to illustrate risks, suggest that random arrangements should be avoided for low-numeracy audiences (and probably low literacy audiences as well, given the correlation between numeracy and literacy). The data also show that interactive graphics can increase the perceived usefulness of risk information and, under certain circumstances, perceived risk and interest in preventive action. The research project should also be expanded to explore the development and evaluation of more complex types of risk graphics. Other types of interactive graphics are possible, and applications that are more immersive, visually appealing, and gamelike might produce even more realistic experiences of risk. One unanswered question is the best way to present comparative information, such as risks with and
without behavior change, or risks with two different treatment options; my findings of poor
accuracy in estimating random stick figure displays suggest that side-by-side random displays
may be a poor choice for this purpose. A second question is the best way to present time-
dependent risk information, such as survival after breast cancer treatment. A third set of
interrelated topics for continued research is the best way to combine risk graphics with numbers
and verbal text, how best to explain graphics with text, and which concepts to illustrate graphically
and which to explain verbally.

Risk graphics, as Figure 8.1 shows, are also a subset of the larger set of medical and
scientific data graphics. Many other types of graphics are relevant to consumer informatics
applications, such as illustrations of personal laboratory values such as serum glucose,
cholesterol, blood pressure, or weight (including trends over time and comparisons to target
goals), medication regimens and other treatment regimens (including the dosage, sequences,
and timing of therapies), nutrition information, and utility information such as personal values
assigned to different health states. Methods similar to the ones used in the current project can be
applied to develop and assess the best visuals. In addition, animation and interactivity could be
used to illustrate other types of data for consumers as well. Interactive animations would seem to
be particularly promising for purposes such as explaining mechanisms of health and disease;
improving understanding of complex scientific concepts such as genetics, evolution, or herd
immunity; or providing instruction for self-management of disease.

Exploring the best types and applications of medical and scientific data graphics for expert
audiences is also a promising direction for future research. The findings in the literature reviews
presented in Chapters 2, 3 and 5 show that experts as well as consumers can be swayed or
confused by the design of graphics. One interesting issue is that even within expert audiences,
the requirements for good graphic design are likely to differ between scientific publications and
informatics applications. Readers of scientific journals may be willing to spend time examining
and parsing a complex data graphic that presents novel and important information. By contrast,
clinical informatics applications must present graphics that are intuitive and immediately
comprehensible without careful examination. Again, the qualitative and quantitative methods used
in this project could be applied to developing and assessing data graphics for these audiences and settings as well. Interactive animations would appear to be particularly useful for allowing clinicians to explore individual patient data, and allowing researchers and policy-makers explore data sets. The "interactive publications" project of the National Library of Medicine, which aims to provide journal readers with the ability to interact with information including published data sets, is one project heading in this direction.

One theme that ties all these potential projects together is the possibility of tailoring graphics for people with different levels of expertise or skills. The current data adds to the literature showing that numeracy (a) varies widely even among well educated people, and (b) strongly affects the interpretation of graphics. The interactive search graphics developed in this project, however, sharply reduced the differences in interpretation and decisions between high- and low-numeracy readers. These or other types of graphics could be used to compensate for low numeracy, thus helping people perform at a higher level. A particularly promising and under-researched area is the topic of tailoring risk graphics, other types of data graphics, and numerical information in general to compensate for low numeracy. There are hints in the literature of the best ways to tailor numerical information for low numeracy readers (e.g., avoiding ratios with 'difficult' denominators such as 1 in 312 in favor of those with intuitive denominators such as 3 in 1000), but there has been no comprehensive development of solutions for different numeracy levels.

A second theme is the importance of designing graphics to support specific goals (i.e., information, education, persuasion, or decisions). Design decisions can have unintended consequences on these goals. For example, alterations in graphic design have frequently been found to increase or decrease the persuasiveness of information (27-29). Ethically, it is acceptable in some situations for communications to be persuasive, and unacceptable in others. The field of public health has long sought to persuade people to avoid hazards and adopt healthy lifestyles, and the potential individual and societal benefits are generally agreed to justify persuasive or manipulative tactics (172, 240). In a free-market economy, persuasion is also accepted in advertising for consumer products. However, public concern about the ethics of
persuasion has resulted in some federal restrictions, such as the U.S. Food and Drug Administration requirement that direct-to-consumer ads for prescription drugs provide information about risks and refrain from exaggerating benefits (241). By contrast, communication in other settings must not be persuasive. For example, persuasion and coercion are unacceptable when recruiting participants for medical research, and consent forms must include unbiased explanations of both risks and benefits (242). Similarly, a patient should choose therapies upon an informed understanding of risks, benefits, and personal values, not upon persuasion by physicians or marketers (243). Data intended for the scientific community, as in scientific publications, should also be presented impartially. These ethical implications warrant consideration in future research into data presentations and graphics.

The ultimate goal of the current project, as well as the future research described here, is to integrate the most effective communication methods into communication and informatics systems for information, education, persuasion (such as tailored risk communication), and decision-making (including shared medical decision-making). Public health is best served by maximizing the persuasive impact of health promotion messages for the public, while maximizing the informational and educational impact of information about research findings and medical options for consumers and experts alike.
LITERATURE CITED


29. Schirillo JA, Stone ER. The greater ability of graphical versus numerical displays to increase risk avoidance involves a common mechanism. Risk Analysis 2005;25(3):555-566.


100. Lowe RK. Interrogation of a dynamic visualization during learning. Learning and Instruction 2004;14(3):257-274.


133. Mayer RE. The promise of multimedia learning: using the same instructional design methods across different media. Learning and Instruction 2003;13(125-139).


177. Epstein RM, Alper BS, Quill TE. Communicating evidence for participatory decision-making. JAMA 2006;291:2359-2366.


203. Lau EW, Ng GA. Visual illusions created by survival curves and the need to avoid potential misinterpretation. Medical Decision Making 2002;22:238-244.


APPENDIX A. DETAILS OF SYSTEMATIC REVIEW PROCESS


To update to the Lipkus and Hollands review of 1999, we searched for evaluative (experimental or focus group) studies of graphs describing probabilities, frequencies, or chances of health events that had not been described by Lipkus and Hollands. We searched the following databases for 1998 and 2005 inclusive.

I. MEDLINE

1. First search: Lipkus and Hollands

   RESULTS of "find citing": n = 11
   HITS:
   • (1) Lee, D. H.; Mehta, M. D. Evaluation of a visual risk communication tool: effects on knowledge and perception of blood transfusion risk. Transfusion. 43(6):779-787, June 2003. (*find citing*, n = 2; *find similar*, n = 0)
   RESULTS of find similar (1999-2005): n = 235
   HITS:
   • dup of Lee and Mehta
   • (2) Fuller R. Dudley N. Blacktop J. Risk communication and older people-understanding of probability and risk information by medical inpatients aged 75 years and older. [Journal Article] Age & Ageing. 30(6):473-6, 2001 Nov. (*find citing n = 4; *find similar n = 0*)

2. General search

   A. MeSH plus keywords

   A1 (MeSH headings) decision making OR data display OR risk taking:
   AND (key words) graph OR health promotion OR risk perception OR risk communication OR pictogram OR numeracy;
   limits: evaluation studies, clinical trial, controlled clinical trial, randomized controlled trial, or validation studies, and human;
   years: 1999-2005
   RESULTS n = 98
   HITS:
   • (3) Zikmund-Fisher BJ. Fagerlin A. Ubel PA. What's time got to do with it? Inattention to duration in interpretation of survival graphs. [Journal Article] Risk Analysis. 25(3):589-95, 2005 Jun. (*find citing n = 0; *find similar n = 0*)
   • (5) Armstrong, Katrina; Weber, Barbara; Ubel, Peter A.; Peters, Nikki; Holmes, John; Schwartz, J Sanford Individualized survival curves improve satisfaction with cancer risk management decisions in women with BRCA1/2 mutations. Journal of Clinical Oncology vol 23 no. 36, 2005; 9319 - 9328.

   A2 data display/ and (decision making/ or decisions.tw) 1998-2006
   n = 39
   HITS:
   • (8) Elting LS. Martin CC. Cantor SB. Rubenstein EB. Influence of data display formats on physician investigators' decisions to stop clinical trials: prospective trial with repeated measures.[see comment]. [Journal Article] BMJ. 318(7197):1527-31, 1999 Jun 5. (*find citing n = 5; *find similar n = 0*)

B. Keywords only

b1. “numeracy” 1998-2005
n = 53
HITS:
  (find citing n = 1; find similar: n = 126)

b2. “graph or graphic or computer graphics” and “risk” 1998-2005
n = 245
HITS:
  duplicate of zikmund fisher et al above.
  duplicate of armstrong et al 2001 above


b3: (probability OR risk) and “pictorial” 1998 -- 2006
n = 38
  dup of other fuller

• (14) *Fuller R. Dudley N. Blacktop J. How informed is consent? Understanding of pictorial and verbal probability information by medical inpatients. [Journal Article] Postgraduate Medical Journal. 78(923):543-4, 2002 Sep.  (Find citing n = 3; find similar n = 0)

b4. “Statistics” and “comprehension”

n = 14, no hits

Total articles found in all MEDLINE searches (i.e., 1 or 2 or 3 or ....) = 969

II. PSYCNINFO
Lipkus and Hollands does not appear in this database, so find similar and find citing are not possible.

1. CONTROLLED terms plus keywords
   (controlled terms) decision making, graphical displays, numbers [numerals], risk taking, visual displays, risk perception
   AND (keywords) graph or chart or pictogram or numeracy
   1998-2005
   limit human
   n = 182
   HITS:
   dup of schapira et al 2004 (find citing n = 1; find similar: n = 0)
   dup of zikmund fisher (find citing n = 0; find similar: n = 0)

2. Keywords only
2a. graphical data 1998-2005 n = 7
   HITS from previous find similar:
   • (16) *Stone, Eric R; Sieck, Winston R; Bull, Benita E; Yates, J. Frank; Parks, Stephanie C; Rush, Carolyn J. Foreground:background salience: Explaining the effects of graphical displays on risk avoidance. [References]. [Journal; Peer Reviewed Journal] Organizational Behavior and Human Decision Processes. Vol 90(1) Jan 2003, 19-36.  (find citing n = 0; find similar: n = 0)

2b. “Statistics” and “comprehension”
n = 28
   • (17) *Parrott R, Silk M, Dorgan K, Condit C, Harris T. Risk comprehension and judgments of statistical evidentiary appeals: when the picture is not worth a thousand words. Human Communication Research. 2005;31(Three):423-452.  (Find citing n = 0; find similar n = 0)

Total for all PsycINFO searches = 245
III. CINAHL
1. Lipkus and Hollands
find citing n = 11
dup of lee and mehta
find similar 1998-2006, limit research, exclude word “safe” and “safety”
n = 422
dup of lee and mehta
dup of fuller et al
2. Keywords:
graphics and risk 1998-2006
n = 9
HITS:
dup of lipkus and hollands
dup of Schapira 2001
3. Controlled terms
(controlled terms) decision making or risk-taking behavior or health promotion
and
(controlled terms) graphics or computer graphics
1998-2006
n = 27
dup of Schapira 2001

Total for CINAHL searches n = 676

IV: ACM portal
1. (risk or probability) and (health or medical)
1999-2006, journal articles
journal articles
n = 23
2. (consumer health informatics) and (risk or probability)
1999-2006, journal articles
n = 3
3. interactive health communication
1999-2006 journal articles
n = 27
4. risk AND (graphical perception or graphical user interface or human computer interaction)
1999-2006 journal articles
n = 110
5. risk AND (graphs or computer graphics)
1999-2006 journal articles
n = 126
6. health AND (decision-making, decisionmaking, decision making)
n = 32 into
7. health AND risk
n = 11

V. Health Psychology web site
fielded search: risk and graph
n= 76

VI: Risk Analysis web site
fielded search: risk and graph
n = 90
dups of zikmund,schirillo, and Shapiro 2004
Risk communication and health
n = 12

VII. Medical Decision making
(risk or probability) and (graph or scale)
2001-2006 (because of web site limitation)
n = 54
dup of Shapiro 2004

**SUMMARY**
The JAMIA article summarizes 16 studies published 1998-2005...
1. Armstrong et al JGIM 2001
2. Armstrong et al Medical Decision making 2002
3. Elting et al 1999
4. Fortin et al 2001
5. Fagerlin et al 2005
6. Fuller et al 2001
7. Fuller et al 2002
8. Feldman_Stewart et al 2000
9. Lee and Mehta 2003
10. Parrott et al 2005
11. Schapira MDM 2001
12. Schapira et al Risk Analysis 2004
13. Schirillo Stone 2005
15. Woloshin et al 2000
17. Royak-Schaler 2004
18. Gurmankin et al 2005
19. Armstrong et al JCO 2005

...plus 5 older articles identified from bibliography searches. Although they were published before 1998, they had not been described in Lipkus and Hollands.
20. Mazur and Hickam 96
21. Mazur and Hickam 94a (MDM)
22. Mazur and Hickam 94b (Theoret Surg)
23. Mazur and Hickam 93b (Journal of Am Geriatric Soc)
24. Mazur and Hickam 1990

In addition, for completeness, the article mentions several studies for context that were cited in Lipkus and Hollands.
1. Stone et al 1997
2. Baty et al 1997
3. Mazur and Hickam, Medical Decision Making,1993
4. Johnson and Slovic 1995
5. Sandman Weinstein Miller 1994
7. Weinstein, Sandman and Hallman 1994
APPENDIX B. STUDY QUESTIONNAIRE

Pictures of people
In this section, we will show you pictures of groups of people, and ask you to guess what percentage of the people are blue. You will have 10 seconds to see each picture before it disappears. Don't worry about being accurate. Don't count the people. Just take a guess!

[practice exercise, followed by 6 pages each containing one graphic]

Health decisions
In this section, we will give you two imaginary stories about health risks. Then you will see several questions about each story. Try to answer the questions as if the story had happened to you.

Each story is illustrated with a picture to give you a feeling for the risks. In the picture, the blue people have the disease, and the yellow people do not.

When you're ready to start the survey, click twice on the picture. [followed by practice exercise for switch and search graphic groups]

Story 1
Dr. Smith tells you that there’s a new disease going around. This disease causes a very high fever and painful headaches for at least one week. Some people have to go to the hospital, and a few of them will die. Your risk of getting this disease is 29%.

He recommends a free vaccine to prevent the disease. It will lower your risk to almost zero. But there’s a 9% chance of a side effect. This side effect is long-term nerve damage, which will make some of your muscles very weak.

To see the questions, please click twice on the picture.

1. Without getting the vaccine, I would feel that I'm going to get the disease this year
   - o agree strongly
   - o agree mildly
   - o disagree mildly
   - o disagree strongly

2. With no vaccine, I would feel very vulnerable to the disease
   - o agree strongly
   - o agree mildly
   - o disagree mildly
   - o disagree strongly

3. If I don’t get the vaccine, I think my chances of getting the disease this year would be
   - o almost zero
   - o very small
   - o small
   - o moderate
   - o large
   - o very large
   - o almost certain

4. If I don’t get the vaccine, I think my chances of getting the disease this year would be ____%.
   (Please enter a number between 0 and 100.)

5. The picture showing the risk of disease is realistic
   - o agree strongly
   - o agree mildly
   - o disagree mildly
   - o disagree strongly

6. The picture is an accurate way of showing the risk of disease
   - o agree strongly
   - o agree mildly
   - o disagree mildly
   - o disagree strongly

7. With this information, I would plan to get the vaccine.
   - o agree strongly
   - o agree mildly
   - o disagree mildly
   - o disagree strongly

8. If I wanted the vaccine, I am sure I could get it.
   - o agree strongly
   - o agree mildly
   - o disagree mildly
   - o disagree strongly

9. I would avoid the vaccine because of the risk of side effects.
   - o agree strongly
   - o agree mildly
   - o disagree mildly
   - o disagree strongly

10. The vaccine is effective at preventing the disease.
    - o agree strongly
    - o agree mildly
    - o disagree mildly
    - o disagree strongly

11. If I got this disease, I would be seriously ill.
    - o agree strongly
    - o agree mildly
    - o disagree mildly
    - o disagree strongly

Now imagine that Dr. Smith gave you some more information.

Story 2
Dr. Smith said your risk of getting heart disease during your lifetime was 6%. Heart disease can lead to chest pain, heart attacks, and other problems.
He recommends a new cholesterol drug to prevent the disease. His clinic will give you the drug for free.

If you take the drug, the risk of heart disease will go down to almost zero. But there's a 2% chance of a side effect. This side effect causes long-term muscle pain and kidney problems.

To see the questions, please click twice on the picture.

[same 11 questions, followed by S-TOFHLA and numeracy questionnaire]

**Final questions**

1. How old are you?___
2. What is your gender?
   - male
   - female
3. What was the last grade of school you completed?
   - did not graduate from high school
   - high school diploma
   - technical school degree
   - some college
   - college degree or graduate degree
4. What is your race or ethnicity? (Check all that apply.)
   - African-American
   - Asian
   - Caucasian or White
   - Hispanic
   - Middle Eastern
   - Native American
   - Pacific Islander
   - other
5. Have you ever had the flu (influenza)?
   - yes
   - no
   - not sure
6. Have you ever been told by a doctor that you have heart disease?
   - yes
   - no
   - not sure
7. Have you ever gotten a side effect from a prescription drug?
   - yes
   - no
   - not sure
8. How would you describe your overall health?
   - very bad
   - poor
   - fair
   - good
   - excellent
9. How often do you use a computer?
   - rarely or never
   - a few times a month, or less
   - a few times a week
   - every day
10. How comfortable do you feel using the computer mouse?
   - not comfortable
   - somewhat comfortable
   - very comfortable

11. Do you have an email address?
   - yes
   - no

Earlier in this survey, you read two stories illustrated with pictures of people. Please rate how you feel about each of these sentences.

12. The pictures helped me understand the risks in the story.
   - agree strongly
   - agree mildly
   - disagree mildly
   - disagree strongly
13. The pictures were confusing.
   - agree strongly
   - agree mildly
   - disagree mildly
   - disagree strongly