# Improving User Performance in Conditional Probability Problems with Computer-Generated Diagrams

Vince Kellen, Susy Chan, and Xiaowen Fang

DePaul University, College of Computing and Digital Media, Chicago, IL, USA
vkellen@uky.edu, {schan,xfang}@depaul.edu

**Abstract.** Many disciplines in everyday life depend on improved performance in probability problems. Most adults struggle with conditional probability problems and prior studies have shown user accuracy is less than 50%. This study examined user performance when aided with computer-generated Venn and Euler-type diagrams in a non-learning context. Following relational complexity, working memory and mental model theories, this study manipulated problem complexity in diagrams and text-only displays. Partially consistent with the study hypotheses, complex visuals outperformed complex text-only displays and simple text-only displays outperformed complex text only displays. However, a significant interaction between users' spatial ability and the use of diagram displays led to a reversal of performance for low-spatial users in one of the diagram displays. Participants with less spatial ability were significantly impaired in their ability to solve problems with less relational complexity when aided by a diagram.

**Keywords:** Human-computer interaction, diagrams, Bayesian reasoning, relational complexity, spatial ability, working memory, individual differences, mental models.

## 1 Objective

Many people struggle with conditional probability problems, which are common in the medical field and in any discipline that needs to communicate and interpret diagnostic evidence [1,2,3]. The research regarding facilitating user performance with computer visuals has yielded conflicting results. The objectives of this study were to a) determine ways of improving user performance with conditional probability problems in non-learning contexts using computer-generated Venn and Euler type diagrams and b) provide a set of design guidelines for designers of systems that have to convey conditional probability problems.

## 2 Background Literature

According to theories of working memory [4,5] and mental models research [6,7,8,9], visual displays should aid in these tasks as they can make better use of working

M. Kurosu (Ed.): Human-Computer Interaction, Part III, HCII 2013, LNCS 8006, pp. 183–192, 2013.

memory, reduce cognitive load and facilitate mental model construction, thus generating better judgments. The conditional probability problem is a common and difficult quantitative reasoning task. While a significant amount of research has examined why people make errors in conditional probability reasoning problems [10,11,12] very little of it has looked systematically at optimizing people's performance using visuals. Moreover, practically no research has applied theories of working memory to guide facilitation of these kinds of problems.

Counter-intuitively, some research is finding that Venn and Euler type diagrams, which try to clearly depict set inclusions and conditional relationships directly, may actually hurt performance [12,13,14]. Mental models theory [6], relational complexity theory [15] and explanations of working memory limits [16,17] can provide guidance for designing visuals to help people solve these problems.

## 3 Conditional Probability Problems

Several studies show participants' accuracy on conditional probability problems is low to intermediate, ranging from 6% of the problems correct to 62% depending on the textual format [1,18]. An example of a conditional probability problem, shown below, has been used in several studies [1]:

> *The probability of breast cancer is 1% for women at age forty who participate in routine screening. If a woman has breast cancer, the probability is 80% that she will get a positive mammography. If a woman does not have breast cancer, the probability is 9.6% that she will also get a positive mammography. A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?*

The normative answer requires applying a form of Bayes' theorem to calculate the posterior probability that a woman with positive results has cancer. The symbols $H$ and $-H$ denote the two hypotheses (cancer and no cancer, respectively) and $D$ denotes the data representing the positive test results. The formula for computing $p(H \mid D)$ is as follows.

$$p(H \mid D) = \frac{p(H)p(D \mid H)}{p(H)p(D \mid H) + p(-H)p(D \mid -H)} = \frac{(.01)(.80)}{(.01)(.80) + (.99)(.096)} = .078$$

Prior studies show that even physicians estimate probabilities incorrectly when presented in text form. Eddy [10] shows that 95% of the physicians queried estimated the probability of p(H | D) in this problem between 70-80% not 7.8%. Other prior research indicates people's accuracy is typically poor for this text-only, probability representation of the problem [1,2,12].

Gigerenzer and Hoffrage [1] have argued that expressing conditional probability problems as natural frequencies rather than as probabilities improves peoples' performance. Natural sampling mimics the process of encountering instances in a population sequentially. Participants should perform better when probability problems are expressed as natural frequencies. A natural frequency representation has a simpler mathematical form of Bayes' theorem:

$$p(H \mid D) = \frac{d\,\&\,h}{d\,\&\,h + d\,\&\,{-h}} = \frac{8}{8+95} = 0.078$$

### 3.1 Use of Visuals in Conditional Probability Problem Solving

A lesser but growing body of research has addressed how to teach people to solve statistical problems [3,11]. Unfortunately, many everyday settings that require statistical reasoning abilities involve people who haven't been taught how to reason with statistics or have failed to retain the associated skills. One common diagram, an Euler diagram, failed to provide significant improvement in either a probability or frequency format [12] and performed worse than iconic displays [13] which depict each item in the frequency of observations as an icon. Iconic displays are superior to Venn diagrams because they better approximate 'actual ecological presentations' [13]. In a study examining deductive reasoning [13], researchers found a disadvantage for Euler diagrams, which performed worse than similar text-only representations.

### 3.2 Mental Models, Working Memory and Reasoning

Both behavioral and neuropsychological mental models theory research [6,7,9] shows that people reason in these problems not by following linguistic rules, but by means of visuospatial models of the problem structure. While verbal processing and the phonological loop would need to be employed to process the text of these problems and for certain math operations solving the problems requires manipulation of the premises and relations within the problem using visuospatial processing and central executive cognitive resources. Several dual task studies show that the VSSP and CE are involved in problem solving, consistent with mental models theory [8,19]. Mental models need not be visual, and can represent non-spatial relationships, such as kinship, or non-visual precepts in deductive reasoning. Model based reasoning interferes with other spatial tasks, not necessarily with concurrent visual tasks without a spatial component [9]. Due to the overlapping use of the VSSP for both reasoning and visual processing, it is possible that use of visual diagrams might be interfering with performance on reasoning problems.

Cowan [20] defines short term memory as "faculties of the human mind that can hold a limited amount of information in a very accessible state temporarily" and surmises that working memory is used to retain partial results while working out a math problem without paper or "to combine the premises in a lengthy rhetorical argument." Baddeley's [4] influential model of working memory holds that verbal-phonological and visuospatial representations are stored separately and managed and manipulated by the central executive. In Baddeley's model [4], working memory is generally viewed as several components working together [20]. Although early research into working memory capacity showed that people can recall about seven chunks of information in short-term memory tasks [21] more recent and extensive research is placing a lower limit of about four chunks on the capacity limits in short-term memory [16,17] across both visual and verbal tasks.

### 3.3 Relational Complexity

Relational complexity is defined as the number of relations that must be processed in parallel, that is, at the same time, to perform a task [15,22] As relational complexity increases, the processing complexity of a cognitive process increases. While a cognitive process may be made up of several steps, the complexity of the process is the measure of the complexity of its most complex task. Process complexity is affected by how many interdependent elements have to be processed in parallel at one time, not how many have to be processed over time. The maximum relational complexity adults can normally process is the quaternary relationship which binds four interactive elements. Planning a correct strategy to solve a problem depends first on representing the complete structure of the problem. Relational theory predicts that people's need to process premises jointly will increase processing load [15]. Mental model construction is constrained by the complexity of the model itself and the working memory capacity, especially the ability for the person to focus attention to the features of the model. In a neurological study [23], researchers confirmed that increasing relational complexity from 0 to 4 in a reasoning problem dramatically increased the percentage of errors and the time in processing, as well as increasing activation in areas of the brain associated with working memory.

## 4 How Visuals Can Improve Performance

Conditional probability problems typically contain three and can contain as many as six or seven elements that might need to be processed in parallel. Based on relational complexity theory, mental models theory, cognitive load theory and the theories of working memory discussed here, for visuals to reliably help with these kinds of problems, designers need to consider the following:

1. Through perceptual cueing mechanisms such as highlighting and controlling what is visible in the display, restrict the user's focus of attention to the task of relating two independent elements at a time.
2. Use visuals to depict relationships between entities rather than describing the relationship with words.
3. Repeat the process of cueing and processing fewer rather than more elements at a time to help the user make a series of correct inferences.

Using this approach, designers ought to be able to improve user performance in making correct inferences by minimize element interactivity in the display. This encourages users to build partial models based on fewer interacting elements at a time and substitute images for words to depict relationships.

## 5 Hypotheses

Based on relational complexity theory, working memory theory and mental models research, diagrams perform better than text and displays with lower element interactivity should perform better than those with high element interactivity. This leads to two hypotheses.

- **Hypothesis 1:** Under both high and low levels of relational complexity, users will demonstrate higher performance with diagrams than with text.

Diagrams should help by further relieving working memory by substituting visuals in place of some text (reducing the amount of text processing needed) and by providing visual cues and guidance than can't be equivalently done in a simpler text-only representation. While performance in the diagram or text-only condition is expected to be strongly influenced by complexity, the role of the visual under these two conditions is of interest.

- **Hypothesis 2:** Under both the text and diagram conditions, users will demonstrate higher performance with problems of lower relational complexity than with problems of higher relational complexity.

Since working memory demands will be reduced with low relational complexity displays that have less element interactivity, users will generate more correct answers. Some of the problems with the use of visuals in prior research may be a result of researchers failing to consider the number of interacting elements in conditional probability problems, the overall working memory demands of these problems and the limited working memory resources available.

## 6 Method

This study recruited 158 participants from a large research university. This study used both students and staff participants with varying ages and backgrounds in order to better approximate real-world populations. This study used a 2X2 factorial design with the two factors being complexity and display type. Complexity has two levels: low (three interacting elements at a time) and high (six interacting elements at a time). Display type has two levels: text-only and diagrams. Because participants were randomly assigned to one of the four groups in the factorial design, this was a between-subjects design. The independent variables were relational complexity (low relational complexity and high relational complexity) and display type (text-only and diagrams). Four treatments were used. Text-only treatments with low and high relational complexity and Venn-type diagram treatments of low and high relational complexity were used and presented via a computer display. Participants expressed their answers in whole numbers and were allowed to round up or down to the next integer. User performance was measured by number of correct answers out of the ten in the battery of problems. Participants had 60 minutes to complete the 10 problems and all participants completed them within the allotted time. To test for individual differences and their effect on problem performance, participants completed a pre-test survey that collects background information (age, gender, educational level, probability reasoning experience, skill, and whether English is a native language). Because spatial ability may contribute to, if not be the primary source of people's performance with regards to relational complexity [24] and is used for processing spatial diagrams [25], this study measured spatial ability [26] and used it as a covariate.

## 7 Results

Approximately one-third of the participants (57) were older than 22 and 28 of the participants were not in college. None of the participant background variables were found to be significantly or marginally associated with performance. Performance was measured by the count of correct problems out of the ten problems presented. Table 1 shows the means, standard deviation and sample size for each condition.

**Table 1.** Performance means

| Complexity | Display Type | | | | | |
|---|---|---|---|---|---|---|
| | *Diagram* | | | *Text* | | |
| | M | SD | n | M | SD | n |
| *Complex* | 5.38 | 2.62 | 40 | 3.32 | 2.00 | 37 |
| *Simple* | 6.39 | 3.60 | 44 | 7.41 | 3.13 | 37 |

Since the hypotheses call for testing for main effect and for interaction with the covariate, a two-way ANCOVA analysis was conducted to test for the influence of complexity and display type and to check for any influence of spatial ability (as measured by the VZ-2 test score). The overall analysis of covariance was highly significant (F(6,151) = 17.56, p<.0001, η2=.41). The analysis of the model effects are shown in Table 2.

**Table 2.** ANCOVA for Performance

| Source | DF | SS | MS | F | Pr > F |
|---|---|---|---|---|---|
| Display type | 1 | 11.47 | 11.47 | 1.76 | 0.1860 |
| Complexity | 1 | 236.97 | 236.97 | 36.47 | <.0001 |
| Spatial ability | 1 | 276.10 | 276.10 | 42.50 | <.0001 |
| Complexity*Display type | 1 | 62.53 | 62.53 | 9.62 | 0.0023 |
| Display type*Spatial ability | 1 | 75.39 | 75.39 | 11.60 | 0.0008 |
| Complexity*Spatial ability | 1 | 18.30 | 18.30 | 2.82 | 0.0953 |

This study showed a significant main effect for complexity but not for display type. However, the main effects (complexity and display type) interacted and the covariate (spatial ability) interacted with Display type. The ANCOVA analysis also showed that the slopes of the regression lines for each treatment are not the same, also confirming an interaction between display type and spatial ability.

Pairwise comparisons to test the hypotheses for performance were conducted with one-tailed t-tests to match the hypotheses, adjusted for the differing covariate means and for multiple inferences. Looking at the first hypothesis, H1, participants in this study demonstrated higher levels of performance with diagrams than with text for only the high relational complexity (complex) treatments (5.38 versus 3.32 correct, t(75)=3.29, p=.0034, Cohen's d=.76). Participants did not demonstrated better performance with diagrams than with text in the low relational complexity (simple) condition (6.39 versus 7.41 correct, t(79)=-1.53, p=.7878). Thus H1 is partially confirmed.

Looking at the second hypothesis (H2), participants demonstrated better performance with low relational complexity (simple) displays over high relational complexity (complex) displays for only the text treatments (7.41 versus 3.32 correct, t(72)=6.57, p<.0001, Cohen's d=1.55.) Participants did not demonstrate improved performance with simple displays than with complex ones for the diagram treatments (6.39 versus 5.38 correct, t(82)=1.46, p=.0840). Thus H2 is partially confirmed. Table 3 lists the comparisons for H1 and H2.

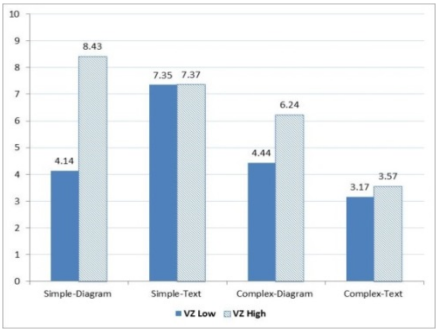
**Table 3.** Pairwise *comparisons*

| | M | SD | M | SD | *t*-test | | Cohen's *d* |
|---|---|---|---|---|---|---|---|
| *H1: Diagrams > Text* | | | | | | | |
| 3D>3T | 6.39 | 3.60 | 7.41 | 3.13 | -1.53 | | |
| 6D>6T | 5.38 | 2.62 | 3.32 | 2.00 | 3.29 | ** | .76 |
| *H2: Low RC > High RC* | | | | | | | |
| 3D>6D | 6.39 | 3.60 | 5.38 | 2.62 | 2.07 | | |
| 3T>6T | 7.41 | 3.13 | 3.32 | 2.00 | 6.54 | *** | 1.55 |

$^{**}p \le .01$ $^{***}p \le .001$

This study finds partial support for both hypotheses. Using diagrams helped participants in the complex condition. Because diagrams enabled use of visual processing and text processing that conserves working memory and provides cues and guidance not possible in a text display, users performed better. Reducing relational complexity also helps improve performance, but for the text-only condition

In looking at the ANCOVA results, diagram displays and spatial ability are affecting user performance. To understand the nature of the main effects interaction it is helpful to see performance by a split of spatial ability (below). Low spatial participants performed 51% worse (4.14 versus 8.44, N=21 and 23 respectively) on the simple diagram and 28% worse (4.44 versus 6.24, N=19 and 21 respectively) on the complex diagram than the high spatial participants. Additional two-tailed t-tests confirmed this finding for both the simple diagram (3D) displays (t(42)=4.79, p<.0001, Cohen's *d*=1.46) and the 6D



(t(38)=2.31, p<.0264, Cohen's *d*=.74). Also of note, low-spatial and high spatial participants did not differ in their performance on both the simple and complex text-only versions of the problem (p=.5662 and p=.9267 respectively). The diagram displays are impairing low-spatial participants, precipitously for the simple diagram (3D) treatment and significantly for the complex diagram (6D) treatment. The text displays performed equally well for high- and low-spatial participants.

## 8 Discussion

While overall, reducing relational complexity improves user performance and more complex problems are aided by diagrams, spatial ability is interacting with display type. Diagram displays hurt the performance of participants with less spatial working memory. Spatial ability may be a shared working memory resource serving both non-visual reasoning and perceptual processes. Perceptual processes, being bottom-up processes, may be taking priority and interfering with the reasoning processes for low-spatial users. In terms of Baddeley's model [4], this may mean that the VSSP may be required for non-visual reasoning processes like those used in conditional probability problems as well as for processing an external representation, which in this case is a diagram. Low-spatial users may have less capacity to process or inhibit and restrict bottom-up processing, thus causing visual diagram processing to conflict with spatial working memory needed to construct and validate a correct mental model. While these bottom-up perceptual processes may be the basis of improvement for high-spatial users as it benefits them in offloading cognitive work to the visual stream, these bottom up processes may be crowding out working memory needed for reasoning in the low-spatial user.

Since spatial ability is also associated with other intelligence measures [27,28], one might also expect that low-spatial participants would have reduced performance on the text-only version of the problem due to presumed deficits in other non-visual cognitive capacities. This study failed to identify that effect. Low and high spatial users performed similarly on text-only versions of the problem. Since this study relied exclusively on the computer presentation of each problem by not allowing any secondary tools such as notepaper to interact with, participants had no choice but to interact with the computer display to solve the problem. This study suggests that capacity limits in spatial working memory may be the culprit for the diminished diagram performance for low-spatial users.

To help users with conditional probability problems in everyday, non-learning contexts, this study recommends the following guidelines:

1. Reduce relational complexity. With the exception of the simple diagram display for low-spatial users, simpler problems helped users produce more correct answers.
2. For low-spatial users, use simpler, text-only displays. These displays will facilitate performance better than diagrams will.
3. For high-spatial users, Venn and Euler-type diagrams can be safely used and can improve performance.

4. Use natural frequencies. All of the problems used in this study used natural frequencies and the best performing treatment, the simpler text-only displays, had a mean of 7.51 out of 10 correct. If all components of the conditional probability problem must be displayed at once, the diagrammatic display will provide the best performance overall.
5. If spatial ability measures are available for the target audience, designers can personalize the display. Simpler diagram displays can be used with high-spatial users and simpler text-only displays can be used with low-spatial users.

## 9 Conclusion

This study has further clarified how diagrams can (or cannot) aid in solving conditional probability problems. Reductions of complexity can improve performance, especially with text displays. Diagrams can help, but they can hurt too. Individual differences in spatial ability matter when processing Venn and Euler-type diagrams. The impairment with diagrams for people with lower spatial ability could be the source of some of the conflicting findings on the use of diagrams to facilitate conditional problem. The extensive collection of research into people's performance in solving conditional probabilities and Bayesian reasoning has not yet applied concepts and frameworks from cognitive psychology like relational complexity and theories of working memory. The recent research into mental models and the neural correlates of working memory and reasoning can provide linkages between the functional and physical descriptions of the mind. These frameworks can be applied to improve our understanding of user performance problems with computer diagrams without having to refer to less testable constructs such as 'frequency coding in the mind' posited by frequentist interpretations of performance [18].

## References

1. Gigerenzer, G., Hoffrage, U.: How to improve Bayesian reasoning without instruction: frequency formats. Psychological Review 102, 684–701 (1995)
2. Kahnamen, D., Lovallo, D.: Timid choices and bold forecasts: a cognitive perspective on risk taking. Management Science 39(1), 17–31 (1993)
3. Sedlmeier, P.: How to improve statistical thinking: Choose the task wisely and learn by doing. Instructional Science 28, 227–262 (2000)
4. Baddeley, A.: Working memory. Clarendon Press, Oxford (1986)
5. Miyake, A., Shah, P.: An Introduction. In: Miyake, A., Shah, P. (eds.) Models of Working Memory. Cambridge University Press (1999)
6. Johnson-Laird, P.N., Legrenzi, P., Girotto, V., Legrenzi, M.S.: Naive probability: A mental model theory of extensional reasoning. Psychological Review 106, 62–88 (1999)
7. Fangmeier, T., Knauff, M., Ruff, C., Sloutsky, V.: fMRI evidence for a three-stage model of deductive reasoning. Journal of Cognitive Neuroscience 18(3), 320–334 (2006)
8. Knauff, M., Jola, C., Strube, G.: Spatial reasoning: No need for visual information. In: Montello, D.R. (ed.) COSIT 2001. LNCS, vol. 2205, pp. 447–457. Springer, Heidelberg (2001)

9. Ruff, C.C., Knauff, M., Fangmeier, T., Spreer, J.: Reasoning and working memory: Common and distinct neuronal processes. Neuropsychologia 41, 1241–1253 (2003)
10. Eddy, D.M.: Probabilistic reasoning in clinical medicine: Problems and opportunities. In: Kahneman, D., Slovic, P., Tversky, A. (eds.) Judgment under Uncertainty: Heuristics and Biases, pp. 249–267. Cambridge University Press, Cambridge (1982)
11. Gigerenzer, G., Hoffrage, U.: Overcoming difficulties in Bayesian Reasoning: A reply to Lewis & Keren and Mellers & McGraw. Psychological Review 106, 425–430 (1999)
12. Sloman, S.A., Over, D., Slovak, L., Stibel, J.M.: Frequency illusions and other fallacies. Organizational Behavior and Human Decision Processes 91, 296–301 (2003)
13. Brase, G.L.: Pictorial representations in statistical reasoning. Applied Cognitive Psychology (2008), doi:10.1002/acp.1460
14. Calvillo, D.P., DeLeeuw, K.E., Revlin, R.: Deduction with Euler Circles: Diagrams That Hurt. In: Barker-Plummer, D., Cox, R., Swoboda, N. (eds.) Diagrams 2006. LNCS (LNAI), vol. 4045, pp. 199–203. Springer, Heidelberg (2006)
15. Halford, G.S., Wilson, W.H., Phillips, S.: Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. Behavioral and Brain Sciences (21), 803–865 (1998)
16. Cowan, N.: The magical number 4 in short-term memory: A reconsideration of mental storage capacity. Behavioral and Brain Sciences 24, 87–185 (2000)
17. Cowan, N., Elliott, E.M., Saults, J.S., Morey, C.C., Mattox, S., Hismjatullina, A., Conway, A.R.: On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. Cognitive Psychology 51, 42–100 (2005)
18. Brase, G.L., Cosmides, L., Tooby, J.: Individuation, counting, and statistical inference: the role of frequency and whole-object representations in judgment under uncertainty. Journal of Experimental Psychology: General 127(1), 3–21 (1998)
19. Klauer, K.C., Stegmaier, R., Meiser, T.: Working memory involvement in propositional and spatial reasoning. Thinking and Reasoning 3(1), 9–47 (1997)
20. Cowan, N.: What are the differences between long-term, short-term, and working memory? In: Sossin, W.S., Lacaille, J.-C., Castellucci, V.F., Belleville, S. (eds.) Progress in Brain Research, vol. 169. Elsevier B.V. (2008)
21. Miller, G.A.: The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review 63, 81–97 (1956)
22. Halford, G.S., Wilson, W.H., Phillips, S.: Relational knowledge: The foundation of higher cognition. Trends in Cognitive Sciences 14(11), 497–505 (2010)
23. Kroger, J.K., Sabb, F.W., Fales, C.I., Bookheimer, S.Y., Cohen, M.S., Holyoak, K.J.: Recruitment of anterior dorsolateral prefrontal cortex in human reasoning: a parametric study of relational complexity. Cerebral Cortex 12, 477–485 (2002)
24. Halford, G.S., Cowan, N., Andrews, G.: Separating cognitive capacity from knowledge: A new hypothesis. Trends in Cognitive Science 11(6), 236–242 (2007)
25. Hegarty, M.: Individual differences in use of diagrams as memory in mechanical reasoning. Learning and Individual Differences 9(1), 19–42 (1997)
26. Ekstrom, R.B., French, J.W., Harman, H.H.: Manual for kit of factor-referenced cognitive tests. Educational Testing Service, Princeton (1976)
27. Hegarty, M., Waller, D.: Individual differences in spatial ability. In: Shah, P., Miyake, A. (eds.) The Cambridge Handbook of Visuospatial Thinking, pp. 121–169. Cambridge University Press, Cambridge (2005)
28. Miyake, A., Friedman, N.P., Rettinger, D.A., Shah, P., Hegarty, M.: How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. Journal of Experimental Psychology: General 130(4), 621–640 (2001)