
Article:

Okan, Y, Galesic, M and Garcia-Retamero, R (2016) How People with Low and High Graph Literacy Process Health Graphs: Evidence from Eye-tracking. Journal of Behavioral Decision Making, 29 (2-3). pp. 271-294. ISSN 0894-3257

https://doi.org/10.1002/bdm.1891

Word count: 8,074

How People with Low and High Graph Literacy Process Health Graphs: Evidence from Eye-Tracking

Yasmina Okan,[1,2] Mirta Galesic,[3] and Rocio Garcia-Retamero [2,3]

[1] Centre for Decision Research, Leeds University Business School, University of Leeds, UK

[2] Department of Experimental Psychology, University of Granada, Spain

[3] Max Planck Institute for Human Development, Center for Adaptive Behavior and Cognition (ABC), Germany

Authors' Note

Corresponding author: Yasmina Okan, Centre for Decision Research, Leeds University Business School, University of Leeds, Maurice Keyworth Building, Leeds, LS2 9JT, United Kingdom; E-mail: y.okan@leeds.ac.uk; Phone: +44 113 343 2622.

This study is part of grants awarded by the Informed Medical Decisions Foundation (United States) to Yasmina Okan (ref. # 0230-1), and to Rocio Garcia-Retamero by the Ministerio de Economía y Competitividad (Spain; ref. # PSI2011-22954 and PSI2014-51842-R), and was supported by the Max Planck Society. The authors declare independence from these funding agencies in each of the following: design of the study; collection, management, analysis, and interpretation of the data; and preparation of the manuscript. This research was conducted as part of the first author's doctoral dissertation at the University of Granada. We would like to thank all members of the committee for their feedback. We are indebted to Andrés Catena, Edward T. Cokely, Merideth Gattis, Jacob Orquin, Rebecca Rhodes, Priti Shah, the editors, and three anonymous reviewers for their valuable comments and suggestions. We are also grateful to Mona Merkert, Gregor Caregnato, and Christian Elsner for their help in conducting the studies and to Anita Todd for editing the manuscript.

Abstract

Graphs facilitate the communication of important quantitative information, often serving as effective decision support tools. Yet, graphs are not equally useful for all individuals, as people differ substantially in their graph literacy—the ability to understand graphically presented information. Although some features of graphs can be interpreted using spatial-to-conceptual mappings that can be established by adults and children with no graphing experience (e.g., “higher bars equal larger quantities”), other features are linked to arbitrary graph conventions (e.g., axis labels and scales). In two experiments, we examined differences in the processes underlying the comprehension of graphs presenting medical information in individuals with low and high graph literacy. Participants’ eye movements were recorded while they interpreted graphs in which information in conventional features was incongruent with that conveyed by spatial features. Results revealed that participants with low graph literacy more often relied on misleading spatial-to-conceptual mappings and misinterpreted the data depicted. Higher graph literacy was often associated with more time spent viewing the conventional features containing essential information for accurate interpretations. This suggests that individuals with high graph literacy are better able to identify the task-relevant information in graphs, and thus attend to the relevant features to a larger extent. Theoretical, methodological, and prescriptive implications for customization of decision-support systems are discussed.

Keywords: Graph comprehension, eye movements, medical decision making, individual differences, graph literacy

How People with Low and High Graph Literacy Process Health Graphs: Evidence from Eye-Tracking

Graphical displays such as line plots, bar charts, and icon arrays can serve as highly valuable tools for overcoming difficulties in the comprehension of numerical concepts, thus serving as highly effective decision support tools (Ancker, Senathirajah, Kukafka, & Starren, 2006; Garcia-Retamero & Cokely, 2011, 2013; Lipkus, 2007). Unfortunately, graphs are not equally useful for all individuals, as people in the general population differ substantially in their ability to understand graphically presented information (Galesic & Garcia-Retamero, 2011; Kutner, Greenberg, Jin, & Paulsen, 2006). These differences can affect the extent to which individuals benefit from visual displays (Gaissmaier et al., 2012; Garcia-Retamero & Cokely, 2014; Garcia-Retamero & Galesic, 2010; Okan, Garcia-Retamero, Cokely, & Maldonado, 2012). Yet the processes underlying graph comprehension in individuals with varying levels of graph literacy are not well understood. We used eye-tracking methodology to investigate this issue.

Individual Differences in Graph Literacy

Graph literacy refers to one’s ability to understand graphically presented information and includes general knowledge about making inferences from different graphic formats (Freedman & Shah, 2002; Shah & Freedman, 2011). Like other types of literacy (e.g., prose and document literacy; Kutner et al., 2006), higher graph literacy is associated with higher educational levels (Galesic & Garcia-Retamero, 2011), highlighting that developing this skill requires knowledge acquired through formal education and experience with graphs. Graph literacy can include mental representations stored in long-term memory that contain knowledge about the properties of different kinds of displays and procedures for interpreting them (i.e., *graph schemas*; Freedman & Shah, 2002; Maichle, 1994; Peebles & Cheng, 2001, 2003; Pinker, 1990; Ratwani & Trafton, 2008; Shah, Freedman, & Vekiri, 2005), which can exert a top-down influence on graph interpretations. Individuals with higher graph literacy may have more complete schemas, which can contribute to recognizing specific types of graphs, identifying the most relevant features in each graph, and making accurate interpretations of the information depicted (Maichle, 1994; Shah & Freedman, 2011). However, there is a dearth of research examining whether individuals with low and high graph literacy differ in the extent to which they engage in the different types of processes necessary to extract information from graphs. Addressing this question is key to achieve a better understanding of the mechanisms driving differences in performance linked to graph literacy, as well as to identify graph design principles that can promote understanding among less graph literate individuals.

Processes Involved in Graph Comprehension

Prominent graph-comprehension models have identified three types of processes in which viewers engage when making inferences from graphical displays, such as line or bar graphs (e.g., Carpenter & Shah, 1998; Kosslyn, 1989; Lohse, 1993; Pinker, 1990; Simkin & Hastie, 1987). The first is *encoding the visual pattern* to identify the principal features in graphs (e.g., lines with different slopes), and it involves making different visual judgments of the elements (e.g., judgments of position along a scale, slope, length, or angle; Cleveland & McGill, 1986).

The second process is the *translation of the identified visual features into conceptual relations*. For example, variations in the size of spatial features (e.g., bars of different heights) can be used to indicate variations in the quantity of the variables represented. Spatial features are those contained in the pattern, including bars of different heights, or lines following an increasing or decreasing trend. There is evidence suggesting that these translations of spatial into conceptual information—*spatial-to-conceptual mappings*—are non-arbitrary and are governed by general cognitive constraints, as certain mappings (e.g., "high equals more," "steeper equals faster") emerge consistently in adults and children with no graphing experience (Gattis, 2002, 2004; see also Gattis & Holyoak, 1996). One of these constraints stems from viewers' experience with their physical environment (Tversky, 2001, 2009). To illustrate, in the physical world, larger quantities of substances typically reach higher positions along the vertical dimension (Lakoff & Johnson, 1980; Tversky, Kugelmass, & Winter, 1991). By applying this real-world experience to graphs, viewers can infer that higher data points represent larger values. Hence, often spatial features (e.g., bars of different heights) can convey meaning independent of viewers' level of graph literacy.

The third process involves *determining the referents of the concepts identified* by associating them with the specific variables shown in the graph and their numerical values (Carpenter & Shah, 1998; Shah & Carpenter, 1995). This process entails identifying and inferring information from *conventional features* in graphs, including the title of the graph, axis labels, legends, and numerical values on the scales, and integrating this information with that extracted in the first two processes. For instance, in line plots or bar graphs it is necessary to identify the variables represented on the *x* and the *y* axes and the values that these variables take. Contrasting with spatial features, conventional features are linked to arbitrary graph conventions and cannot be interpreted directly on the basis of real-world experience (Okan, Garcia-Retamero, Galesic, & Cokely, 2012). Viewers with low graph literacy may be less likely to have schemas including arbitrary graph conventions. Thus, they might be less prone to identify the relevant conventional features for accurate interpretation of a given graph and to incorporate this information in their interpretations. In contrast, highly graph literate individuals may more readily direct their attention to labels or scales that contain information required to understand a graph correctly.

Conflicts in Graphs and the Role of Conventional Features

The relevance of identifying and inferring information from conventional features can vary depending on specific properties of the graphical displays. If information conveyed by spatial features (e.g., bar heights) is congruent with that conveyed by conventional features, viewers could neglect conventional features and nevertheless reach correct interpretations by relying on spatial-to-conceptual mappings. However, if such congruency does not exist, identifying and inferring information from conventional features becomes critical to reach a correct interpretation. This can occur when spatial features of the graph convey a different meaning from textual information in the title and axis labels (*textual conflicts*) or numerical values on the scale (*scale conflicts*). For instance, a graph with a textual conflict might present the percentage of people *without* different types of allergy (as indicated in the title and axis label), implying that higher bars do not represent more prevalent allergies. In these cases, recognizing such a conflict and taking into account information in conventional features is crucial to override spatial-to-conceptual mappings and avoid misinterpretations.

Differences in the accuracy of understanding graphs with conflicts could arise from at least two mechanisms, which are linked to the third process of graph comprehension outlined above (i.e., determining the referents of the identified concepts). One possibility is that people with low and high graph literacy differ in the extent to which they attend to the relevant conventional features (i.e., those that are critical for accurate understanding of a given graph). This is in line with the information reduction framework proposed by Haider and Frensch (1996, 1999), which suggests that more skilled individuals have acquired the ability to distinguish between task-relevant and task-redundant information, and allocate more attention to the former. Accordingly, more graph literate individuals may be better able to identify and attend to the task-relevant information in graphs (here, the relevant conventional features). This greater allocation of attention would increase their likelihood of detecting a conflict in a graph and interpreting it correctly, as compared to less graph literate individuals.

An alternative possibility is that differences in performance linked to graph literacy stem primarily from differences in conceptual understanding about the meanings of elements of graphs, and mental operations on such elements. That is, individuals with low and high graph literacy might allocate a similar amount of attention to the relevant conventional features, but less graph literate ones may fail to incorporate this information at a conceptual level in their interpretations (for a related distinction in terms of perceptual and conceptual stages, see Haider & Frensch, 1999).

The first possibility outlined above should be reflected in longer viewing times of relevant conventional features for participants with high graph literacy, as compared to those with low graph literacy. However, longer viewing times may stem from allocating more attention to such features both before and after a conflict is detected. More graph literate individuals might initially allocate more attention to the relevant features than less graph literate ones, due to the ability of the former group to identify task-relevant information in graphs. In addition, once the conflict is noticed, both more and less graph literate individuals may allocate further attention to the relevant conventional features, in order to examine the nature of the conflict. Thus, to understand when the differences in viewing times linked to graph literacy occur, it is necessary to consider whether participants detected the conflict in a given graph (i.e., whether they interpreted the graph correctly or not).[1]

If more graph literate participants initially allocate more attention to the relevant conventional features, their viewing times for such features should be longer than those of less graph literate ones, even for graphs in which both groups failed to detect the conflict. In contrast, in the infrequent cases when both more and less graph literate participants detect the conflict, initial attention to relevant conventional features will likely be similar for both groups, as accurate interpretations of graphs with conflicts cannot be reached without attention to such features. Following conflict detection, allocation of further attention may be larger among participants with high graph literacy, in which case viewing times in correct trials may also be longer among such individuals. Alternatively, the allocation of further attention may not vary depending on graph literacy, resulting in similar overall viewing times in correct trials for both groups.

Finally, as noted above, differences in performance linked to graph literacy may instead stem primarily from differences in conceptual understanding. If this were the case, then viewing times of the relevant conventional features should not vary as a function of graph literacy, regardless of whether graphs are interpreted correctly or not.

To determine if and how graph literacy predicts the amount of attention that viewers allocate to critical conventional features, we conducted two experiments in which we recorded the eye movements of participants with low and high graph literacy. Recordings took place while participants interpreted line graphs and bar graphs displaying quantitative medical information (i.e., prevalence of different diseases or effects linked to different treatments). In both experiments we included a set of graphs constructed in such a way that following spatial-to-conceptual mappings grounded in experience with the environment would lead to erroneous interpretations. In Experiment 2 we also included a set of graphs where following spatial-to-conceptual mappings would lead to correct interpretations.

Experiment 1

Experiment 1 included four graphs containing conflicts between spatial and conventional features. Taking into account the roles of prior knowledge outlined above, we proposed two hypotheses. First, in line with recent findings reported by Okan et al. (2012b), we predicted that low graph literacy would be associated with a stronger tendency to interpret graphs on the basis of spatial-to-conceptual mappings. For graphs containing conflicts, this should be reflected in a larger proportion of incorrect responses corresponding to mappings among participants with low graph literacy (e.g., they might often assume that the highest bar represents the highest value) (H1).

Second, in line with the information reduction framework proposed by Haider and Frensch (1996, 1999), we expected that individuals with low graph literacy would be less likely to recognize and attend to the conventional features that are essential to reach a correct interpretation, according to the conflict present in the graph. This should be reflected in relatively longer times spent viewing such features for participants with high graph literacy, as compared to those with low graph literacy, in particular when graphs are interpreted incorrectly (H2).

Method

Participants

A total of 52 participants were recruited from the respondent pool of the Max Planck Institute for Human Development in Berlin. Technical problems prevented recording the eye movements of four participants. The final sample consisted of 48 participants (50% female), mean age of 25.7 years (*SD* = 3.3, range 19–34 years), 52% with up to high school education and 48% with at least some college. Participants were paid 10 Euros for taking part in the study.

Materials

Eye-tracking equipment. Participants’ eye movements were recorded by a Tobii T120 Eye Tracker. To define fixations we used the built-in fixation filter available in Tobii Studio (v. 2.0.3) with a fixation radius of 30 pixels on a screen with a resolution of 1,280 × 1,024 pixels. For all analyses we took into account fixations that lasted at least 100 ms, as this decreases noise in the data (Peebles & Cheng, 2003).

Stimuli. We constructed four graphs presenting medical information, such as prevalence of different diseases and effects linked to different treatments. In two of the graphs, essential information was included in the numerical scale for the *y* axis (graphs with *y*-axis-scale conflicts; see graphs G1 and G2 in Appendix A); the other two graphs contained essential information in the title and in the textual label for the *y* axis (graphs with textual conflicts; see graphs G11 and G12 in Appendix A). To illustrate, one of the graphs involving a scale conflict was a line graph presenting data about the percentage of people with a fictitious disease. The numerical scale on the *y* axis was inverted (i.e., values increased from top to bottom; see graph G1 in Appendix A). Participants were asked to find the year in which the percentage of people with the disease was highest. To answer this question correctly, participants had to attend to the scale to infer that the usual spatial-to-conceptual correspondence between height and quantity was reversed. An example of the graphs involving textual conflicts is a bar graph presenting data about percentages of people *without* a fictitious disease in different clinics (see graph G11 in Appendix A). Participants were asked to identify the clinic in which the percentage of people with the disease was highest. To answer this question correctly, participants had to attend to the title and the label for the *y* axis to infer that the usual spatial-to-conceptual mapping was reversed (i.e., they had to infer that higher bars represented lower values). All materials were implemented as a Web questionnaire using the platform Unipark (www.unipark.de).

Coding of eye fixations. For each graph we defined a set of areas of interest (AOIs) corresponding to the conventional features containing essential information to answer the question in each case, according to the types of conflicts present (i.e., titles, *y* axes labels and scales). AOIs encompassed the relevant conventional features in each case (see Appendix B for details on the size of AOIs). For each participant, we computed the total time spent viewing (fixating on) each of the AOIs. The total time spent viewing each AOI and the number of fixations were highly correlated (mean correlation = .95 across all variables computed), so for the sake of simplicity we report only the former.

Measurement of graph literacy. Graph literacy was measured using the scale developed by Galesic and Garcia-Retamero (2011). This scale consists of 13 items dealing with the communication of medical risks, treatment efficiency, and prevalence of diseases, and covers four frequently used graph types—line plots, bar charts, pies, and icon arrays. Because the scale was designed for the general population, to achieve better differentiation of graph literacy in our somewhat better educated sample, we also included four more difficult items from other scales.[2] The total score for each participant was computed by adding the score for these four items to the score obtained in the scale developed by Galesic and Garcia-Retamero (2011). For some analyses, we split participants into two groups according to the median graph literacy score for the total sample (i.e., 14.5 of the total 17). Participants with low graph literacy (*n* = 24) answered on average 12.5 (*SD* = 1.6) items correctly, while participants with high graph literacy (*n* = 24) answered on average 16.2 (*SD* = 0.8) items correctly.

Measurement of numeracy. We also assessed participants' numeracy skills (i.e., the ability to understand and manipulate different numerical expressions of probability; Lipkus, Samsa, & Rimer, 2001; Peters, 2012). Participants' numeracy was measured using the 11 items included in the general and expanded numeracy scales developed by Lipkus et al. (2001). The correlation of graph literacy with numeracy was .38.

Procedure

The experiment took on average 23.2 min (*SD* = 5.7) to complete. After a standardized calibration exercise, participants were presented with the four graphs depicting medical information. Afterwards, they completed the items measuring graph literacy and numeracy, and answered some demographic questions. The study was approved by the Ethics Committee of the Max Planck Institute for Human Development.

Data Analyses Overview

First, we examined the percentages of correct and incorrect responses. For each participant, we computed the percentage of items in which he or she had provided the incorrect response corresponding to the spatial-to-conceptual mapping (*mapping response,* e.g., assuming that the highest value is the one represented by the highest bar), for each type of conflict (i.e., textual vs. *y*-axis-scale conflict). Analyses of variance (ANOVA) were used to examine the effect of graph literacy and type of conflict on the tendency to show mapping responses, whereas linear mixed models were used to examine the effect of these factors on viewing times of the relevant conventional features. As distributions of time spent viewing the different AOIs were skewed right, we log-transformed the data before conducting the analyses (see Ratcliff, 1993; Yan & Tourangeau, 2008, for a justification). Type of conflict was included as a repeated variable in the mixed models reported below, and a diagonal variance-covariance matrix was used, taking into account the Akaike Information Criterion (Field, 2009). All models contained a random intercept for subjects.[3] Finally, the method of estimation was restricted maximum likelihood in all cases. Statistical analyses were conducted with the Statistical Package for the Social Sciences (SPSS for Windows, version 20.0), and linear mixed models were computed using the MIXED procedure (see, e.g., Janssen, 2012). Results remained unchanged when numeracy was included as a covariate in the analyses reported below.

Results

How Does Graph Literacy Relate to Interpretations of Graphs With Conflicts?

The average proportion of correct responses to the questions across graphs was 56% (*SE* = 10.6), while the average proportion of responses that were both incorrect and consistent with a spatial-to-conceptual mapping (mapping responses) was 37% (*SE* = 9.5). As expected, the average proportion of incorrect responses that were not related to the mapping was low (7%; *SE* = 1.9), indicating that the majority of participants who misinterpreted the graphs did so on the basis of direct spatial-to-conceptual mappings.

The average percentage of mapping responses among participants with low graph literacy was 42% (*SE* = 7.2) for *y*-axis-conflict graphs and 56% (*SE* = 6.9) for textual-conflict graphs. In contrast, participants with high graph literacy showed on average 27% (*SE* = 5.1) mapping responses for *y*-axis-conflict graphs and 23% (*SE* = 6.0) for textual-conflict graphs. A 2 × 2 ANOVA with graph literacy as between-subjects factor and type of conflict as within-subject factor on the average percentage of mapping responses revealed a main effect of graph literacy, $F(1,46) = 15.38$, $p = .001$, supporting H1. All other effects were not reliable (Fs < 2, ps > .1).

How Does Graph Literacy Relate to the Viewing Time of Relevant Conventional Features?

To address this question, we computed a linear mixed model for viewing times of the relevant conventional features, including graph literacy and type of conflict as fixed factors, as well as the interaction between these factors. Both graph literacy and type of conflict predicted viewing times of the relevant conventional features, $F(1,58.09) = 4.45$, $p = .039$, and $F(1,116.34) = 50.50$, $p < .001$, respectively. The interaction was not reliable ($F < 1$, $p >$ .6). Table 1 shows raw and log-transformed mean viewing times for the different areas of the graph. As can be seen in the table, participants with high graph literacy spent more time fixating on the areas containing essential information in each graph, supporting H2. In addition, viewing times were higher for the textual elements (i.e., *y*-axis label and title) than for the *y*-axis scale. When the total time that participants spent viewing the graphs was included as the dependent variable in the model described above both main effects and the interaction were not reliable (*F*s < 2, *p*s > .2).

<Insert Table 1 about here>

To further examine the origin of differences in viewing times between more and less graph literate participants, we added correctness of responding (i.e., whether each item was answered correctly or incorrectly) as a fixed factor to the model described above, as well as two-way and three-way interaction effects. This analysis revealed a main effect of type of conflict, $F(1,114.22) = 66.54$, $p < .001$, a main effect of correctness, $F(1,133.67) = 18.05$, $p < .001$, and an interaction between type of conflict and correctness, $F(1,130.77) = 8.66$, $p = .004$. All other interactions, including the one between graph literacy and correctness, were unreliable (*F*s <1, *p*s > .3). As can be seen in Figure 1, viewing times were longer for correct trials than for incorrect trials, although these differences were reliable only for graphs with *y*-axis-scale conflicts. Participants with high graph literacy had longer viewing times independently of the type of conflict and whether they answered the items correctly. However, unlike in the model without correctness, in this model the main effect of graph literacy was not reliable, $F(1,46.89) = 1.50$, $p = .23$. Figure 1 also shows that more graph literate participants were more likely to give correct answers than less graph literate participants (as reflected in sizes of outer circles, whose radius is proportional to the percentage of incorrect and correct responses, summing to 100% for each type of conflict and graph literacy group).

<Insert Figure 1 about here>

Finally, to further explore more general patterns of eye fixations, we performed additional analyses examining the different types of transitions between different regions, following Carpenter and Shah (1998). Results of these further analyses can be found in Appendix C.

Discussion

The results of Experiment 1 showed that people tend to make erroneous inferences indicating an overreliance on spatial-to-conceptual mappings. In line with our predictions, the tendency to rely on such mappings was larger for less graph literate individuals. As compared to such individuals, more graph literate ones spent more time fixating on the relevant conventional features in graphs with both *y*-axis scale and textual conflicts, regardless of whether graphs were interpreted correctly or not. These findings are in line with the information reduction framework proposed by Haider and Frensch (1996, 1999), according to which skill acquisition leads to differences in attention allocation, with more skilled individuals focusing more on task-relevant information. Although in the task of graph comprehension all regions of a graph are arguably relevant for an accurate interpretation, when graphs contain the conflicts described above some conventional features (e.g., axis labels, values on scales) become particularly relevant. Importantly, the total time that participants spent viewing the graphs did not vary reliably as a function of graph literacy. This suggests that individuals with high graph literacy do not merely engage in a more thorough exploration of all regions of the graphs, but instead allocate more attention to those regions containing the most relevant information for the task at hand.

The findings of Experiment 1 thus support the notion that lower levels of graph literacy are associated not just with a failure to understand and integrate information in key conventional features at a conceptual level, but also with a tendency to spend less time encoding such features. The larger allocation of attention to the relevant conventional features among more graph literate participants likely helped these individuals to more often avoid misinterpretations. Our results also revealed that differences in viewing times linked to graph literacy did not vary depending on whether conflicts were noticed (resulting in a correct answer) or not (resulting in an incorrect answer). However, after including correctness as a factor in our model, the effect of graph literacy on viewing times failed to reach conventional levels of significance, despite the consistent trends towards longer viewing times for more graph literate participants (see Figure 1). We suspected that this was a result of a lack of sufficient statistical power linked to the small number of graphs employed in Experiment 1, or to noise due to baseline individual differences in viewing times. We sought to address these issues in Experiment 2, where we included a larger number of items, including items that did not contain conflicts to determine participants' baseline viewing times.

Experiment 2

Experiment 2 was designed to address three new issues. First, we sought to determine whether the findings observed in Experiment 1 would generalize to a more diverse set of graphs and types of conflict. To this end, we expanded our set of stimuli to include a wider range of graphs with textual conflicts and *y*-axis-scale conflicts, as well as new graphs containing essential information in the *x*-axis scale (i.e., *x*-axis-scale conflicts). Second, as Experiment 1 included only graphs containing conflicts, in Experiment 2 we constructed an equivalent graph without conflict, for each of the graphs with a conflict. This enabled us to determine the extent to which the inclusion of conflicts affected interpretations and viewing times. In addition, these nonconflict graphs enabled us to determine individual baseline viewing times and to control for them when analyzing conflict graphs, in order to reduce noise due to initial individual differences in viewing times. Finally, in Experiment 2, we also aimed to exclude potential confounding factors of the effect of graph literacy, including careless responding, participants' knowledge that graphs can be misleading, and medicine-related knowledge.

We hypothesized that differences in accuracy of understanding linked to graph literacy would be smaller for graphs without conflicts than for graphs with conflicts (H1a).

The reason is that in the former type of graphs, conventional features and spatial features point to the same (correct) interpretation. In contrast, and in line with Experiment 1, for graphs with conflicts we expected that participants with low graph literacy would more often make erroneous interpretations corresponding to spatial-to-conceptual mappings (H1b). We also hypothesized that times spent viewing conventional features in graphs without conflicts would not vary as a function of graph literacy, as such features did not contain essential information for accurate interpretations (H2a). In contrast, for graphs with conflicts we expected that participants with high graph literacy would spend a longer time than those with low graph literacy viewing the relevant conventional features in each case (H2b). As in Experiment 1, we also examined whether any differences in viewing times linked to graph literacy interacted with correctness of responses.

Method

Participants

Ninety-one participants from the database of the Max Planck Institute for Human Development in Berlin were prescreened with the graph literacy scale and the four additional items used in Experiment 1. Their graph literacy scores ranged from 10 to 17, with a mean of 14.5 (*SD* = 1.6). One week after the prescreening, we re-invited 38 participants in the top and bottom quartiles (i.e., with scores ranging from 10 to 13, and from 16 to 17). Due to the limited number of such participants, we also invited another 13 randomly selected participants with scores from 14 to 15. Thus, the final sample included 51 participants (61% female), with a mean age of 25.3 years (*SD* = 4.7, range 18–38 years), 49% with up to high school education and 51% with at least some college. The group of participants with low graph literacy included those who obtained 14 or fewer correct responses (*n* = 24, mean score 12.7, *SD* = .9); the group of participants with high graph literacy included those who obtained

15 or more correct responses (*n* = 29, mean score 16.0, *SD* = .8). Participants were paid 10 Euros for taking part in the study.

Materials

Eye-tracking equipment. The eye-tracking equipment was identical to that used in Experiment 1, and fixations were determined using the same procedure.

Stimuli. In addition to the four graphs constructed in Experiment 1, we constructed 12 new graphs presenting medical information that contained conflicts. In six of the graphs essential information was included in the numerical scale for the *y*-axis (Graphs G1 to G6). In four graphs essential information was included in the numerical scale for the *x*-axis (Graphs G7 to G10). Finally, six graphs contained essential information in the title and in the textual label for the *y*-axis (Graphs G11 to G16). For each of the graphs with conflicts we constructed an equivalent graph without conflict, resulting in a total of 32 graphs. Description of all items with conflicts is given in Table 2, and all graphs (with and without conflicts) can be seen in Appendix A.

<Insert Table 2 about here>

Coding of eye fixations. The same AOIs as in Experiment 1 were defined (i.e., titles, labels for the *y* axes, and scales on the *y* axes). Additionally, we defined an AOI corresponding to the scales on the *x* axes. All AOIs were defined for graphs with and without conflicts (see Appendix B for details on the size of AOIs). As in Experiment 1, the number of fixations and the total viewing times on each AOI were highly correlated (mean correlation = .94 across all variables computed for graphs with conflicts, and .93 for graphs without conflicts).

Measurement of graph literacy. Graph literacy was measured using the same items as in Experiment 1.

Measurement of numeracy. In addition to the numeracy scale used in Experiment 1 (Lipkus et al., 2001), we administered the Berlin Numeracy Test (Cokely, Galesic, Schulz, Ghazal, & Garcia-Retamero, 2012) as it has better psychometric properties and is suitable for educated samples.

Measurement of content knowledge. Participants' medicine-related knowledge was measured using a version of the Minimum Medical Knowledge (MMK) questionnaire (Bachmann et al., 2007) adapted for inclusion in our computerized task.

Measurement of knowledge that graphs can be misleading and careless responses. To measure participants' knowledge that graphs can be misleading we included six items developed by the current authors. Three items concerned graphs in general and three items focused on the medical domain (see Appendix D for more details). Cronbach's alpha for the six items was .81. To identify careless responses we administered the self-report participant engagement items developed by Meade and Craig (2012).[4]

Procedure

The experiment took on average 42 min (*SD* = 7.2) to complete and included three sections. In the first section participants signed a consent form and successfully completed a standardized calibration exercise. They were then presented with the 16 graphs without conflicts. In the second section, participants were presented with the 16 graphs with conflicts. In the third section, participants completed (1) the Berlin Numeracy Test (Cokely et al., 2012), (2) the numeracy scale developed by Lipkus et al. (2001), (3) the items assessing knowledge that graphs can be misleading, (4) the MMK questionnaire, (5) demographic questions, and (6) the items to identify careless responses. As calibration can decrease in accuracy over time, respondents were recalibrated at the beginning of each new section. All remaining aspects of the procedure were identical to that of Experiment 1. The study was approved by the Ethics Committee of the Max Planck Institute for Human Development.

Data Analyses Overview

First, we examined the correlations between graph literacy, numeracy, knowledge that graphs can be misleading, MMK, and careless responding. Next, we examined the percentages of correct and incorrect responses. We conducted ANOVAs to examine how graph literacy relates to interpretations of graphs with and without conflicts, as well as to the tendency to show mapping responses, for graphs with conflicts. Finally, we computed linear mixed models to examine viewing times of conventional features. As in Experiment 1 distributions were skewed right, and viewing times were log-transformed before conducting analyses. When presence of conflict or type of conflict were included as factors in the models described below, these were included as repeated variables, and a diagonal variance-covariance matrix was used. All models contained a random intercept for subjects. Results remained unchanged when numeracy, MMK, knowledge that graphs can be misleading, and careless responding were included as covariates in the analyses reported.

Results

Are Effects of Graph Literacy Confounded by Other Skills, Knowledge, and Motivational Factors?

The correlation of graph literacy with numeracy measured with the Berlin Numeracy Test (Cokely et al., 2012) was .33 ($p = .020$), while it was .32 ($p = .023$) with the Lipkus et al. (2001) numeracy scale. This indicates that even though some of the same abilities might contribute to both graph literacy and numeracy, the amount of shared variance is relatively small. The correlation of graph literacy with MMK was −.09 ($p = .508$), indicating that no linear relationship existed between these variables. Finally, the correlations of graph literacy with knowledge that graphs can be misleading and with scales measuring careless responding developed by Meade and Craig (2012; Diligence, Interest, Effort, and Attention) ranged from

–.15 to .05 (*p*s > .3), suggesting that the effects of graph literacy are unlikely to be confounded with these factors.

How Does Graph Literacy Relate to Interpretations of Graphs With and Without Conflicts?

Table 3 shows the percentage of respondents who gave correct responses to the graphs, as a function of graph literacy. A 2 × 2 ANOVA with graph literacy as between-subjects factor and the presence of conflict as within-subject factor, on the average percentage of correct responses, revealed a main effect of graph literacy, $F(1,49) = 11.22$, $p = .002$, a main effect of the presence of conflict, $F(1,49) = 256.25$, $p = .001$, and an interaction between graph literacy and presence of conflict, $F(1,49) = 4.23$, $p = .045$. As can be seen in Table 3, graphs with conflicts had significantly lower rates of correct responses, as compared to their equivalent versions without conflicts. Overall, the percentage of correct responses was higher for participants with high graph literacy. However, the difference in accuracy between participants with high and low graph literacy was smaller for graphs without conflicts, $d = .59$, than for graphs with conflicts, $d = .80$, supporting H1a.

<Insert Table 3 about here>

Next, we examined the different types of responses provided for graphs with conflicts. The average proportion of correct responses to the questions across graphs was 42% (*SE* = 4.5), while the average proportion of mapping responses was 55% (*SE* = 4.7). As in Experiment 1, the average proportion of incorrect responses that were not related to the mapping was low (4%; *SE* = 0.9).

The average percentage of mapping responses among participants with low graph literacy was 70% (*SE* = 4.7) for graphs with *y*-axis-scale conflict, 48% (*SE* = 7.0) for graphs with textual conflict, and 80% (*SE* = 5.2) for graphs with *x*-axis-scale conflict. In contrast, participants with high graph literacy showed on average 55% (*SE* = 5.4) mapping responses for graphs with *y*-axis-scale conflict, 35% (*SE* = 5.8) for graphs with textual conflict, and 57% (*SE* = 6.8) for graphs with *x*-axis-scale conflict. A 2 × 3 ANOVA with graph literacy as between-subjects factor and type of conflict as within-subject factor on the average percentage of mapping responses revealed a main effect of graph literacy, $F(1,49) = 6.35$, $p = .015$, supporting H1b. This analysis also revealed a main effect of type of conflict, $F(2,98) = 20.74$, $p = .001$, indicating that the percentage of mapping responses was lower for textual-conflict graphs ($M = 41.4$, $SE = 4.5$), as compared to *y*-axis and *x*-axis-conflict graphs ($M = 62.1$, $SE = 3.8$, and $M = 68.8$, $SE = 4.5$). Indeed, conflicts linked to both *y*-axis and *x*-axis scales often remained unnoticed, as most of the graphs with such conflicts were associated with accuracy rates below 50% even among participants with high graph literacy (see Table 3). The interaction between type of conflict and graph literacy was not reliable ($F < 1$, $p > .4$).

How Does Graph Literacy Relate to the Viewing Time of Relevant Conventional Features?

To address this question, we first computed a linear mixed model for viewing times of conventional features, including graph literacy, presence of conflict, and the interaction between graph literacy and presence of conflict as fixed factors. Results revealed a main effect of the presence of conflict, $F(1,923.38) = 187.71$, $p < .001$, and an interaction between graph literacy and presence of conflict, $F(1, 923.38) = 4.46$, $p = .035$. The main effect of graph literacy was not reliable ($F < 1$, $p > .8$). Viewing times for graphs without conflicts were similar for participants with low and high graph literacy, in line with H2a (for raw viewing times, $M = 1.4$ s, $SE = 0.1$ and $M = 1.6$ s, $SE = 0.2$, respectively). In contrast, viewing times were longer among participants with higher graph literacy than among less graph literate participants (for raw viewing times, $M = 3.4$ s, $SE = 0.3$ vs. $M = 2.9$ s, $SE = 0.3$, respectively). Table 4 shows the mean viewing times for the different areas of interest for graphs both with and without conflicts, as a function of the type of conflict and graph literacy.[5]

Next, we performed analyses only for graphs with conflicts. To control for baseline individual variability in viewing times, for each graph and each participant we expressed the viewing time of the relevant conventional features as percentage change from the average time the participant spent viewing the relevant conventional features in graphs without conflicts.[6] In this way we obtained, for each individual separately, a relative increase or decrease in viewing times due to conflicts in each of the graphs. We computed a linear mixed model for these baseline-adjusted times, including graph literacy and type of conflict as fixed factors, as well as the interaction between these factors. This analysis revealed a main effect of type of conflict, $F(2,110.83) = 5.79$, $p = .004$, reflecting that viewing times were higher for the textual elements (i.e., *y*-axis label and title) than for *y*-axis and *x*-axis scales, in line with Experiment 1 (see Table 4). The main effect of graph literacy and the interaction between graph literacy and type of conflict were not reliable (*F*s < 2, *p*s > .1).

<Insert Table 4 about here>

Finally, following the same rationale as in Experiment 1, we added correctness of responding as a fixed factor in the model, as well as the remaining two-way and three-way interactions. This analysis revealed a main effect of type of conflict, $F(2,158.20) = 5.53$, $p = .005$, and a main effect of correctness, $F(1,372.17) = 6.24$, $p = .013$, as well as interactions between correctness and graph literacy, $F(1,372.17) = 9.23$, $p = .003$, and between correctness, graph literacy, and type of conflict, $F(2,175.09) = 4.25$, $p = .016$. There was also an interaction approaching conventional levels of significance between correctness and type of conflict, $F(2,175.09) = 3.05$, $p = .051$. As in the previous model, the main effect of graph literacy and the interaction between graph literacy and conflict were not reliable (*F*s < 2, *p*s > .1).

As can be seen in Figure 2, when responses were correct we observed no reliable differences in viewing times overall between participants with low and high graph literacy. Interestingly, however, less graph literate participants spent more time viewing textual elements than more graph literate ones. In contrast, when responses were incorrect, participants with high graph literacy viewed conventional features longer than less graph literate ones. These results support H2b, as they suggest that individuals with high graph literacy spend more time viewing the relevant conventional features than less graph literate ones even if conflicts are not detected. In addition, viewing times were longer for correct trials than for incorrect trials among less graph literate participants, while this was not the case for more graph literate ones.

<Insert Figure 2 about here>

As in Experiment 1, we also performed additional analyses examining the different types of transitions between global regions (see Appendix C).

Discussion

In Experiment 2 we partially replicated and extended the findings of Experiment 1. Using a more diverse set of graphs containing different types of conflicts, we found that people with lower graph literacy often failed to detect conflicts between information conveyed by spatial features and that conveyed by conventional features, frequently relying on spatial-to-conceptual mappings in their interpretations. In contrast, people with higher graph literacy were more likely to use information from conventional features to override the mappings leading to erroneous conclusions.

Eye-tracking recordings in Experiment 2 showed that participants with high graph literacy spent more time viewing the relevant conventional features, in incorrect trials. These results indicate that more graph literate individuals initially allocated more attention to such features than less graph literate individuals, for graphs with conflicts. In contrast, for correct trials viewing times did not differ as a function of graph literacy, or in some cases were larger among participants with low graph literacy (i.e., for graphs with textual conflicts). This suggests that, when participants with low graph literacy attended the relevant conventional features long enough to detect the presence of a conflict in a given graph, their further attention to these features was similar to that of more graph literate participants. Yet, participants with low graph literacy were often unable to detect conflicts in graphs (as indicated by the sizes of circles in Figure 2), in line with Experiment 1. Moreover, as noted above we also found that less graph literate participants exhibited longer viewing times of textual elements than more graph literate ones, in correct trials. A plausible interpretation is that less graph literate participants needed more time to further process the information conveyed in textual elements, in order to reach a correct interpretation. The fact that interactions between graph literacy and correctness were observed only in Experiment 2 may be attributable to the larger number of items used (as Experiment 1 included only two graphs per conflict type, limiting generalizability of the results), as well as to the reduction in noise due to baseline individual differences in viewing times accomplished in this experiment.

Experiment 2 also included graphs that did not contain any conflicts. For those graphs we found high rates of correct answers for participants with both high and low graph literacy. In line with our predictions, graph literacy did not affect times spent viewing conventional features in graphs without conflicts. This can be accounted for by the fact that for such graphs attending to these features is not crucial to reach a correct interpretation. Thus, for graphs without conflicts participants with low and high graph literacy could rely on information conveyed by spatial features in graphs to reach a correct interpretation.

Finally, in Experiment 2 we excluded a number of possible confounds of the relationship of graph literacy and graph processing. Participants with low graph literacy were not merely more careless, as suggested by the lack of differences linked to graph literacy in items measuring participant engagement, and by the fact that no reliable differences existed in the overall time spent exploring graphs or viewing the questions assessing interpretations. Additionally, graph literacy and medicine-related knowledge were not associated, and participants with low graph literacy knew as well as those with high graph literacy that graphs can be plotted misleadingly. An interesting question for future research would be to examine how graph literacy relates to other factors including working memory limitations (Huestegge & Philipp, 2011; Peebles & Cheng, 2001, 2003), individual differences in spatial abilities (Feeney, Adams, Webber, & Ewbank, 2004), and math anxiety (Ashcraft & Kirk, 2001).

General Discussion

In two experiments we found that participants often failed to detect conflicts between information conveyed by spatial features of graphs (such as slope of a line or height of bars) and conventional features (such as axis labels and scales). This tendency was more common among individuals with low graph literacy, who misinterpreted graphs more frequently than those with high graph literacy. Eye movement data revealed that higher graph literacy was often associated with more time spent viewing the conventional features containing essential information for avoiding misinterpretations.

Theoretical Implications

The present findings expand previous research on perceptual and cognitive processes in graph comprehension (Carpenter & Shah, 1998; Kosslyn, 1989; Lohse, 1993; Pinker, 1990; Simkin & Hastie, 1987), documenting the existence of differences in these processes that are linked to individual differences in graph literacy. Eye-fixation patterns in our study suggest that higher graph literacy is associated with a stronger tendency to direct attention to conventional features containing essential information to detect and override conflicts in graphs. This finding is in accord with the information reduction framework proposed by

Haider and Frensch (1996, 1999), which suggests that successful skill acquisition is characterized by the ability to recognize and focus on task-relevant information (see Orquin, Bagger, & Loose, 2013, for related arguments). In contrast, no reliable differences existed in viewing times of conventional features as a function of graph literacy for graphs without conflicts. This finding is also in line with the information reduction framework because in such graphs attending to conventional features is not crucial to reach a correct interpretation.

Our results are also in line with studies documenting a widespread tendency to interpret graphs on the basis of non-arbitrary spatial-to-conceptual mappings that emerge consistently in adults and children with no graphing experience (e.g., “higher equals more”; Gattis, 2002, 2004). When spatial features are incongruent with the relationships that the graph is supposed to show (e.g., when higher bars do not indicate larger quantities), relying on such mappings can lead to errors in interpretation. Notably, our findings demonstrate that less graph literate individuals show a tendency toward basing their interpretations of graphs primarily on such translations and are thus more likely to misinterpret graphs with conflicts.

Implications for the Graphic Communication of Quantitative Information

Graphs that are available to the public often include misleading characteristics similar to those manipulated in the present study, such as improperly scaled axes (Beattie & Jones, 1992, 2002; Cooper, Schriger, Wallace, Mikulich, & Wilkes, 2003) or longer bars representing lower values (Kosslyn, 2006). To illustrate, Cooper et al. (2003) found that, in a sample of 74 graphs in pharmaceutical advertisements in medical journals, almost 40% contained misleading features, including improperly scaled or split axes and improper baselines. In line with Okan et al.’s (2012b) findings, our results suggest that such graphs are likely to be misinterpreted by people lacking graph literacy skills and point to an important principle for designing graphs that are suitable even for people with low graph literacy: Spatial and conventional features should convey the same meaning. For some graphs, this could help less graph literate people to reach the correct interpretation even without attending to the conventional features.

In addition, methods could be developed to direct attention to essential information in conventional features and increase the likelihood that this information will be encoded and integrated, particularly among less graph literate individuals. For instance, people could be presented with interactive displays that require using mouse clicks to uncover the different regions (see, e.g., Okan, Garcia-Retamero, Cokely, & Maldonado, 2015). Encouraging people to uncover key conventional features in a first step could help them identify referents of the concepts that will be depicted before they make direct spatial-to-conceptual mappings.

Conclusion

The present research adds to the increasing body of work that has successfully employed eye-tracking to achieve a better understanding of the processes underlying the comprehension of different kinds of graphical displays (e.g., Carpenter & Shah, 1998; Huestegge & Philipp, 2011; Kim, Lombardino, Cowles, & Altmann, 2014; Peebles & Cheng, 2001, 2003; Woller-Carter et al., 2012). Here we have identified individual differences in graph comprehension processes linked to viewers' level of graph literacy, shedding light on the mechanisms underlying links between this skill and performance. This work complements and extends work using other process tracing techniques such as verbal protocols, which has provided valuable insights on individual differences in graph reading behaviors (Mason et al., 2014). Our work also illustrates the value of eye-tracking methods to inform the design of improved displays to facilitate the communication of health-relevant quantitative information that could affect potentially life changing decisions (see also Hess, Visschers, Siegrist, & Keller, 2011; Keller, 2011; Keller, Kreuzmair, Leins-Hess, & Siegrist, 2014). In sum, the current findings can play a central role in the development of custom-tailored decision support systems built to inoculate professionals, policy makers, and the general public against potentially distorted and misleading communication.

References

Ancker, J. S., Senathirajah, Y., Kukafka, R., & Starren, J. B. (2006). Design features of graphs in health risk communication: A systematic review. *Journal of the American Medical Informatics Association*, *13*, 608–618.

Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of experimental psychology: General*, *130*, 224.

Bachmann, L. M., Gutzwiller, F. S., Puhan, M. A., Steurer, J., Steurer-Stey, C., & Gigerenzer, G. (2007). Do citizens have minimum medical knowledge? A survey. *BMC Medicine*, *31*, 5–14.

Beattie, V., & Jones, M. J. (1992). The use and abuse of graphs in annual reports: A theoretical framework and an empirical study. *Accounting and Business Research*, *22*, 291–303.

Beattie, V., & Jones, M. J. (2002). Measurement distortion of graphs in corporate reports: An experimental study. *Accounting, Auditing & Accountability Journal*, *15*, 546–564.

Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied*, *4*, 75–100.

Cleveland, W. S., & McGill, R. (1986). An experiment in graphical perception. *International Journal of Man-Machine Studies*, *25*, 491–500.

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin Numeracy Test. *Judgment and Decision Making*, *7*, 25–47.

Cokely, E. T., & Kelley, C. M. (2009). Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. *Judgment and Decision Making*, *4*, 20–33.

Cooper, R. J., Schriger, D. L., Wallace, R. C., Mikulich, V. J., & Wilkes, M. S. (2003). The quantity and quality of scientific graphs in pharmaceutical advertisements. *Journal of General Internal Medicine*, *18*, 294–297.

Feeney, A., Adams, J., Webber, L. & Ewbank, M. (2004). Individual differences in graphical reasoning. *Lecture Notes in Computer Science*, *2980*, 271–285.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London, England: SAGE publications.

Freedman, E. G., & Shah, P. (2002). Toward a model of knowledge-based graph comprehension. In M. Hegarty, B. Meyer, & N. H. Narayanan (Eds.), *Diagrammatic representation and inference* (pp. 59–141). Berlin, Germany: Springer.

Gaissmaier, W., Wegwarth, O., Skopec, D., Müller, A. S., Broschinski, S., & Politi, M. C. (2012). Numbers can be worth a thousand pictures: Individual differences in understanding graphical and numerical representations of health-related information. *Health Psychology*, *31*, 288–296.

Galesic, M., & Garcia-Retamero, R. (2011).Graph literacy: A cross-cultural comparison. *Medical Decision Making*, *31*, 444–457.

Garcia-Retamero, R., & Cokely, E. T. (2011). Effective communication of risks to young adults: Using message framing and visual aids to increase condom use and STD screening. *Journal of Experimental Psychology: Applied*, *17*, 270–287.

Garcia-Retamero, R., & Cokely, E. T. (2013). Communicating health risks with visual aids. *Current Directions in Psychological Science*, *22*, 392–399.

Garcia-Retamero, R., & Cokely, E. T. (2014). The influence of skills, message frame, and visual aids on prevention of sexually transmitted diseases. *Journal of Behavioral Decision Making*, *27*, 179–189.

Garcia-Retamero, R., & Galesic, M. (2010).Who profits from visual aids: Overcoming challenges in people's understanding of risks. *Social Science & Medicine*, *70*, 1019–1025.

Gattis, M. (2002). Structure mapping in spatial reasoning. *Cognitive Development*, *17*, 1157–1183.

Gattis, M. (2004). Mapping relational structure in spatial reasoning. *Cognitive Science*, *28*, 589–610.

Gattis, M., & Holyoak, K. J. (1996). Mapping conceptual to spatial relations in visual reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 231–239.

Haider, H., & Frensch, P. A. (1996). The role of information reduction in skill acquisition. *Cognitive Psychology*, *30*, 304–337.

Haider, H., & Frensch, P. A. (1999). Eye movement during skill acquisition: More evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 172–190.

Hayes, A. F. (2013). *An introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford Press.

Hess, R., Visschers, V. H. M., Siegrist, M., & Keller, C. (2011). How do people perceive graphical risk communication? The role of subjective numeracy. *Journal of Risk Research*, *14*, 47–61.

Huestegge, L., & Philipp, A. M. (2011).Effects of spatial compatibility on integration processes in graph comprehension. *Attention, Perception, &Psychophysics*, *73*, 1903–1915.

Janssen, D. P. (2012). Twice random, once mixed: Applying mixed models to simultaneously analyze random effects of language and participants. *Behavior Research Methods*, *44*, 232–247.

Keller, C. (2011). Using a familiar risk comparison within a risk ladder to improve risk understanding by low numerates: A study of visual attention. *Risk Analysis*, *31*, 1043–1054.

Keller, C., Kreuzmair, C., Leins-Hess, R., & Siegrist, M. (2014). Numeric and graphic risk information processing of high and low numerates in the intuitive and deliberative decision modes: An eye-tracker study. *Judgment and Decision Making*, *9*, 420–432.

Kim, S., Lombardino, L. J., Cowles, W., & Altmann, L. J. (2014). Investigating graph comprehension in students with dyslexia: An eye tracking study. *Research in Developmental Disabilities*, *35*, 1609–1622.

Kosslyn, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology*, *3*, 185–226.

Kosslyn, S. M. (2006). *Graph design for the eye and mind*. New York, NY: Oxford University Press.

Kramarski, B., & Mevarech, Z. R. (2003). Enhancing mathematical reasoning in the classroom: The effects of cooperative learning and metacognitive training. *American Educational Research Journal*, *40*, 281–310.

Kutner, M. A., Greenberg, E., Jin, Y., & Paulsen, C. (2006). *The health literacy of America's adults: Results from the 2003 National Assessment of Adult Literacy*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: University of Chicago Press.

Lipkus, I. M. (2007). Numeric, verbal, and visual formats of convening health risks: Suggested best practices and future recommendations. *Medical Decision Making*, *27*, 695–713.

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, *21*, 37–44.

Lohse, G. L. (1993). A cognitive model for understanding graphical perception. *Human-Computer Interaction*, *8*, 353–388.

Maichle, U. (1994). Cognitive processes in understanding line graphs. In W. Schnotz & R. W. Kulhavy (Eds.), *Comprehension of graphics* (pp. 207–227). Amsterdam, The Netherlands: North-Holland Elsevier Science.

Mason, D., Boase, S., Marteau, T., Kinmonth, A. L., Dahm, T., Minorikawa, N., & Sutton, S. (2014). One-week recall of health risk information and individual differences in attention to bar charts. *Health, Risk & Society*, *16*, 136–153.

Meade, A. W., & Craig, S. B. (2012).Identifying careless responses in survey data. *Psychological Methods*, *17*, 437–455.

Okan, Y., Garcia-Retamero, R., Cokely, E. T., & Maldonado, A. (2012a). Individual differences in graph literacy: Overcoming denominator neglect in risk comprehension. *Journal of Behavioral Decision Making*, *25*, 390–401.

Okan, Y., Garcia-Retamero, R., Cokely, E. T., & Maldonado, A. (2015). Improving risk understanding across ability levels: Encouraging active processing with dynamic icon arrays. *Journal of Experimental Psychology: Applied*. Advance online publication. doi: 10.1037/xap0000045

Okan, Y., Garcia-Retamero, R., Galesic, M., & Cokely, E. T. (2012b).When higher bars are not larger quantities: On individual differences in the use of spatial information in graph comprehension. *Spatial Cognition & Computation*, *12*, 1–25.

Orquin, J. L., Bagger, M. P., & Mueller Loose, S. (2013). Learning affects top down and bottom up modulation of eye movements in decision making. *Judgment and Decision Making*, *8*, 700–716.

Peebles, D., & Cheng, P. C. H. (2001). Graph-based reasoning: From task analysis to cognitive explanation. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the Twenty Third Annual Conference of the Cognitive Science Society* (pp. 762–767). Mahwah, NJ: Erlbaum.

Peebles, D., & Cheng, P. C. H. (2003). Modeling the effect of task and graphical representation on response latency in a graph reading task. *Human Factors*, *45*, 28–45.

Peters, E. (2012). Beyond comprehension: The role of numeracy in judgments and decisions. *Current Directions in Psychological Science*, *21*, 31–35.

Pinker, S. (1990).A theory of graph comprehension. In R. Freedle (Ed.), *Artificial intelligence and the future of testing* (pp. 73–126). Hillsdale, NJ: Erlbaum.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*, 510–532.

Ratwani, R. M., & Trafton, J. G. (2008). Shedding light on the graph schema: Perceptual features versus invariant structure. *Psychonomic Bulletin & Review*, *15*, 757–762.

Shah, P., & Carpenter, P. A. (1995). Conceptual limitations in comprehending line graphs. *Journal of Experimental Psychology: General*, *124*, 43–61.

Shah, P., & Freedman, E. G. (2011). Bar and line graph comprehension: An interaction of top-down and bottom-up processes. *Topics in Cognitive Science*, *3*, 560–578.

Shah, P., Freedman, E. G., & Vekiri, I. (2005).The comprehension of quantitative information in graphical displays. In P. Shah & A. Miyake (Eds.), *The Cambridge* *handbook of visuospatial thinking* (pp. 426–476). New York, NY: Cambridge University Press.

Simkin, D., & Hastie, R. (1987).An information-processing analysis of graph perception. *Journal of the American Statistical Association*, *82*, 454–465.

Tuijnman, A. (2000). *Benchmarking adult literacy in America: An international comparative study (International Adult Literacy Survey)*. Washington, DC: U.S. Department of Education, Office of Vocational and Adult Education.

Tversky, B. (2001). Spatial schemas in depictions. In M. Gattis (Ed.), *Spatial schemas and abstract thought* (pp. 79–112). Cambridge, MA: MIT Press.

Tversky, B. (2009). Spatial cognition: Embodied and situated. In P. Robbins & M. Aydede (Eds.), *The Cambridge handbook of situated cognition* (pp. 201–217). Cambridge, England: Cambridge University Press.

Tversky, B., Kugelmass, S., & Winter, A. (1991). Cross-cultural and developmental trends in graphic productions. *Cognitive Psychology*, *23*, 515–557.

Woller-Carter, M., Okan, Y., Cokely, E. T., & Garcia-Retamero, R. (2012).Communicating and distorting risks with graphs: An eye-tracking study. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 56,* 1723–1727.

Yan, T., & Tourangeau, R. (2008). Fast times and easy questions: The effects of age, experience and question complexity on web survey response times. *Applied Cognitive Psychology*, *22*, 51–68.

Footnotes

[1] We thank the editors for offering this insight, as well as for their valuable suggestions concerning the approach to analyses of viewing times described below.

[2] We included one item from the Kramarski and Mevarech (2003) Graph Interpretation Test, two items assessing graph comprehension from the International Adult Literacy Survey (Tuijnman, 2000), and one item from the National Assessment of Adult Literacy (Kutner et al., 2006).

[3] The random effect of subjects was significant in all models reported in this experiment and in Experiment 2.

[4] We used eight items selected from the Diligence subscale and the six items in the Interest subscale. Item 7 from the Diligence subscale was excluded as it was not applicable to the context of the current study. We included a response scale ranging from 1 (*Completely disagree*) to 4 (*Completely agree*). Cronbach's alpha was .82 for the Diligence subscale and .75 for the Interest subscale. We also included the three single-item measures developed by Meade and Craig evaluating effort expended on the study (Effort), attention to the study (Attention), and whether the respondent felt his or her data should be used for analysis (Use Me), all adapted for the context of our study. The item Use Me was not included in analyses, as all participants provided the same response (i.e., "yes"). All items were translated into German by a native German speaker with excellent knowledge of English and were reviewed by two of the authors.

[5] The total time that participant spent viewing the graphs did not vary as a function of graph literacy, in line with Experiment 1 ($F < 1$, $p > .9$). However, participants with high graph literacy spent significantly longer overall viewing graphs with conflicts ($M = 11.1$, $SE = 0.9$) than graphs without conflicts ($M = 8.6$, $SE = 0.5$), while this was not the case among less graph literate participants ($M = 10.2$, $SE = 1.0$ and $M = 10.0$, $SE = 0.6$ for graphs with and

without conflicts respectively). This finding suggests that graph literacy is associated with strategic differences in allocation of attention, as highly graph literate individuals seem to identify the need to process graphs more thoroughly when they contain conflicts (see also Cokely & Kelley, 2009; Woller-Carter, Okan, Cokely, & Garcia-Retamero, 2012). No reliable differences were observed as a function of graph literacy in the time that participants spent viewing the questions assessing interpretations.

[6] More formally, for each participant s and item i. $adjTC_{si} = (TC_{si} - \overline{TNC_s})/\overline{TNC_s}$, where $adjTC_{si}$ is baseline adjusted viewing time for s^{th} participant and i^{th} item with conflict, TC_{si} is raw viewing time for the same participant and item, and $\overline{TNC_s}$ is the average time that participant s spent viewing items without conflict.

.

Table 1

Raw and Log-Transformed Mean Times Spent Viewing the Relevant Areas of the Graphs and Total Viewing Times in Experiment 1, as a Function of Type of Conflict and Graph Literacy (SEM in Parentheses).

Area of graph	Graphs with *y*-axis-scale conflict		Graphs with textual conflict	
	Low graph literacy	High graph literacy	Low graph literacy	High graph literacy
y-axis scale				
Time	**2.59**	**3.10**	1.15	1.47
	(0.49)	**(0.46)**	(0.17)	(0.24)
Log time	**0.29**	**0.69**	-0.21	0.00
	(0.21)	**(0.19)**	(0.15)	(0.16)
Title &*y*-axis label				
Time	4.34	5.24	**4.96**	**6.44**
	(0.67)	(0.63)	**(0.65)**	**(0.64)**
Log time	1.10	1.38	**1.31**	**1.68**
	(0.15)	(0.12)	**(0.16)**	**(0.10)**
Graph total				
Time	14.98	15.43	12.97	15.49
	(1.46)	(1.39)	(1.05)	(1.39)
Log time	2.58	2.59	2.47	2.62
	(0.09)	(0.09)	(0.08)	(0.09)

Note: Relevant conventional features for each type of conflict are marked in bold. ‘Graph total’ refers to the total time spent viewing all regions contained within the framework of the graph, including the pattern, title, axes scales and labels, and legends (where applicable), and did not include the time spent viewing the questions about the data.

Table 2

Summary of the Experimental Materials

Item	Task
y-axis-scale conflict	
G1. Line graph with inverted scale (values increase from top to bottom), *x* axis below	Find the year in which the percentage of people with a disease was highest
G2. Bar graph with scale with negative values	Identify the therapy that resulted in the *lowest* change in the percentage of people with the disease
G3. Stacked bar graph with excised scale	Identify the ethnic group for which the proportion of people with one type of flu was larger than the proportion of people with another kind of flu
G4. Bar graph with scale with both positive and negative values and bars rising from lower *x* axis. Zero baseline not indicated.	Identify the treatment that resulted in the smallest change in patients' body weight
G5. Line graph with logarithmic scale	Find the year in which the difference between the number of men and women dying after suffering an infection was larger
G6. Line graph with inverted scale (values increase from top to bottom), *x* axis above	Find the age at which the recovery time from a disease was lowest
x-axis-scale conflict	
G7. Bar graph with inverted scale (values increasing from right to left)	Identify the pill that resulted in an increase in the values of a hormone over time
G8. Line graph with inverted scale (values increasing from right to left)	Identify the disease for which the number of affected people increased over time
G9. Bar graph with values not placed at proportional distances	Find the month after which patients' blood iron levels started to increase more slowly
G10. Line graph with values not placed at proportional distances	Find the week after which pain scores started to decrease more slowly
Textual conflict	
G11. Bar graph showing percentages of people *without* a disease	Identify the clinic in which the percentage of people *with* the disease was highest
G12. Bar graph showing the *change in the percentage* of people with different types of cancer during the previous year	Identify the type of cancer that affected the smallest *percentage* of people during the previous year
G13. Bar graph showing the number of *patients per doctor* in different countries	Identify the country that had the highest number of *doctors per patient*

Item	Task
G14. Line graph showing the percentage of people testing *negative* for a disease at different ages	Identify the age at which the percentage of people *diagnosed* with the disease was highest
G15. Line graph showing the number of *patients per nurse* in different years	Find the year in which the number of *nurses per patient* was lowest
G16. Line graph showing the percentage of people who *died* after different weeks of having been exposed to a virus	Find the week in which the percentage of people who *survived* after being expose to the virus was lowest

Note: Graphs G1, G2, G11, and G12 with conflicts were used in Experiment 1. All graphs were used in Experiment 2.

Table 3

Percentage of Respondents Who Gave Correct Responses to the Graphs in Experiment 2, as a Function of Whether They Contained Conflicts or Not, and Graph Literacy

Item	Low graph literacy		High graph literacy	
	Nonconflict	Conflict	Nonconflict	Conflict
y-axis-scale conflict				
G1. Inverted scale, *x* axis below	100%	27%	100%	45%
G2. Scale with negative values	86%	64%	97%	83%
G3. Excised scale	100%	5%	100%	21%
G4. Scale not indicating zero baseline	95%	14%	100%	45%
G5. Logarithmic scale	100%	0%	100%	28%
G6. Inverted scale, *x* axis above	100%	50%	100%	52%
x-axis-scale conflict				
G7. Inverted scale (bar)	91%	27%	97%	38%
G8. Inverted scale (line)	100%	18%	97%	38%
G9. Values not placed at proportional distances (bar)	95%	14%	97%	48%
G10. Values not placed at proportional distances (line)	100%	9%	100%	34%
Textual conflict				
G11. People without disease	100%	45%	100%	55%
G12. Change in the percentage of people with different types of cancer	82%	32%	100%	59%
G13. Number of patients per doctor	100%	50%	100%	59%
G14. People testing negative	95%	45%	100%	66%
G15. Number of patients per nurse	91%	36%	93%	48%
G16. People who died after virus exposure	73%	68%	90%	76%
Average overall	94%	32%	98%	50%

Table 4

Raw and Log-Transformed Mean Times Spent Viewing the Relevant Areas of the Graphs and Total Viewing Times in Experiment 2, as a Function of Whether They Contained Conflicts or Not, Type of Conflict, and Graph Literacy (SEM in Parentheses).

	Graphs with *y*-axis-scale conflict				Graphs with *x*-axis-scale conflict				Graphs with textual conflict			
	Low graph literacy		High graph literacy		Low graph literacy		High graph literacy		Low graph literacy		High graph literacy	
Area of graph	Nonconflict	Conflict	Nonconflict	Conflict	Nonconflict	Conflict	Nonconflict	Conflict	Nonconflict	Conflict	Nonconflict	Conflict
y-axis scale												
Time	**1.01**	**2.18**	**0.55**	**2.53**	1.34	1.02	1.06	1.39	1.01	1.05	0.64	1.36
	(0.20)	**(0.40)**	**(0.07)**	**(0.40)**	(0.22)	(0.20)	(0.18)	(0.29)	(0.14)	(0.15)	(0.07)	(0.20)
Log time	**-0.74**	**-0.34**	**-1.20**	**-0.17**	-0.48	-0.76	-0.61	-0.77	-0.55	-0.53	-0.84	-0.47
	(0.14)	**(0.23)**	**(0.17)**	**(0.24)**	(0.19)	(0.26)	(0.19)	(0.22)	(0.15)	(0.17)	(0.12)	(0.18)
x-axis scale												
Time	1.41	1.55	1.27	1.67	**0.67**	**1.14**	**0.79**	**2.26**	1.35	1.33	1.36	1.40
	(0.13)	(0.13)	(0.07)	(0.12)	**(0.07)**	**(0.39)**	**(0.10)**	**(0.40)**	(0.13)	(0.13)	(0.10)	(0.11)
Log time	0.10	0.12	0.01	0.23	**-0.96**	**-1.05**	**-0.80**	**-0.53**	0.03	0.00	-0.01	0.10
	(0.10)	(0.10)	(0.07)	(0.08)	**(0.13)**	**(0.28)**	**(0.13)**	**(0.27)**	(0.11)	(0.11)	(0.08)	(0.08)
Title &*y*-axis label												
Time	2.10	2.33	2.13	2.59	2.26	1.44	2.30	1.54	**2.94**	**4.92**	**2.71**	**5.00**
	(0.23)	(0.20)	(0.18)	(0.27)	(0.28)	(0.16)	(0.26)	(0.22)	**(0.35)**	**(0.65)**	**(0.23)**	**(0.53)**
Log time	0.28	0.40	0.38	0.55	0.10	-0.31	0.26	-0.09	**0.65**	**1.12**	**0.75**	**1.25**
	(0.13)	(0.14)	(0.11)	(0.13)	(0.15)	(0.20)	(0.15)	(0.16)	**(0.18)**	**(0.19)**	**(0.09)**	**(0.13)**
Graph total												
Time	9.46	10.14	8.16	11.12	12.66	9.51	11.18	11.10	8.65	10.72	7.38	11.15
	(0.90)	(0.85)	(0.48)	(1.01)	(0.94)	(1.09)	(0.83)	(1.37)	(0.78)	(1.08)	(0.50)	(1.03)
Log time	2.07	2.41	1.99	2.41	2.30	2.01	2.25	2.10	1.97	2.13	1.86	2.20
	(0.09)	(0.09)	(0.06)	(0.08)	(0.07)	(0.11)	(0.07)	(0.12)	(0.10)	(0.11)	(0.06)	(0.10)

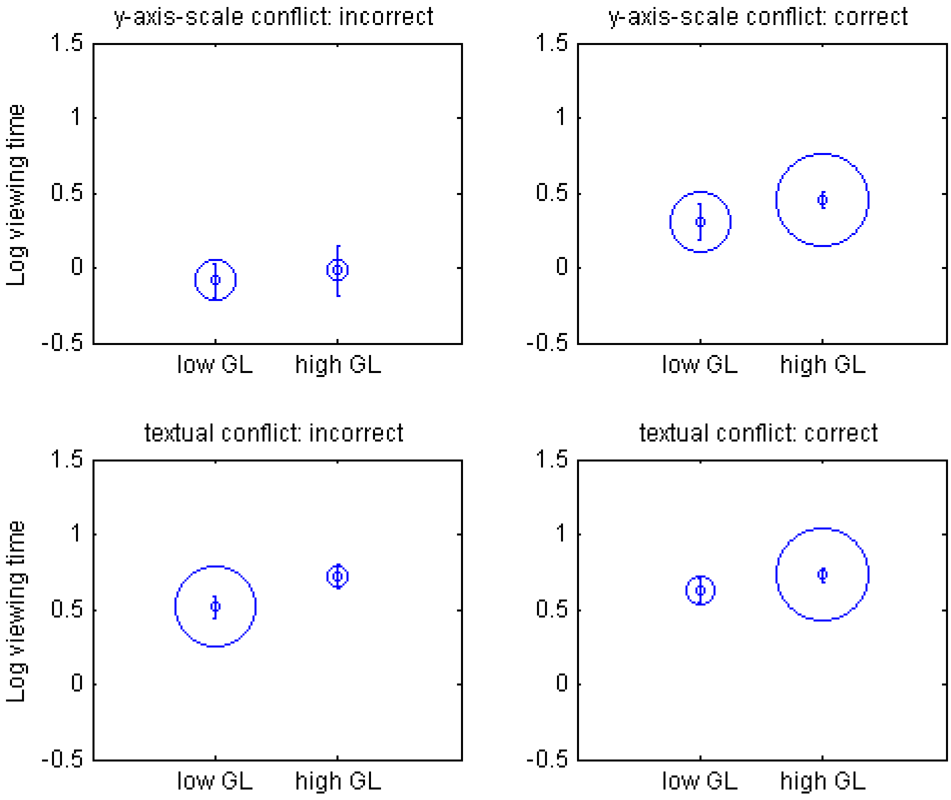
Note: Relevant conventional features for each type of conflict are marked in bold. ‘Graph total’ refers to the total time spent viewing all regions contained within the framework of the graph, including the pattern, title, axes scales and labels, and legends (where applicable), and did not include the time spent viewing the questions about the data.

Figure captions

Figure 1. Average viewing times of less (low GL) and more (high GL) graph literate participants, by type of conflict (*y*-axis-scale or textual) and correctness of response (incorrect or correct). Viewing times for *y*-axis conflict graphs correspond to *y*-axis scales, whereas viewing times for textual conflict graphs correspond to titles and *y*-axis labels. Inner circles denote average log-transformed viewing times. Bars denote +/-1 standard error. Areas of outer circles are proportional to the percentage of incorrect and correct responses for each graph literacy group and type of conflict.

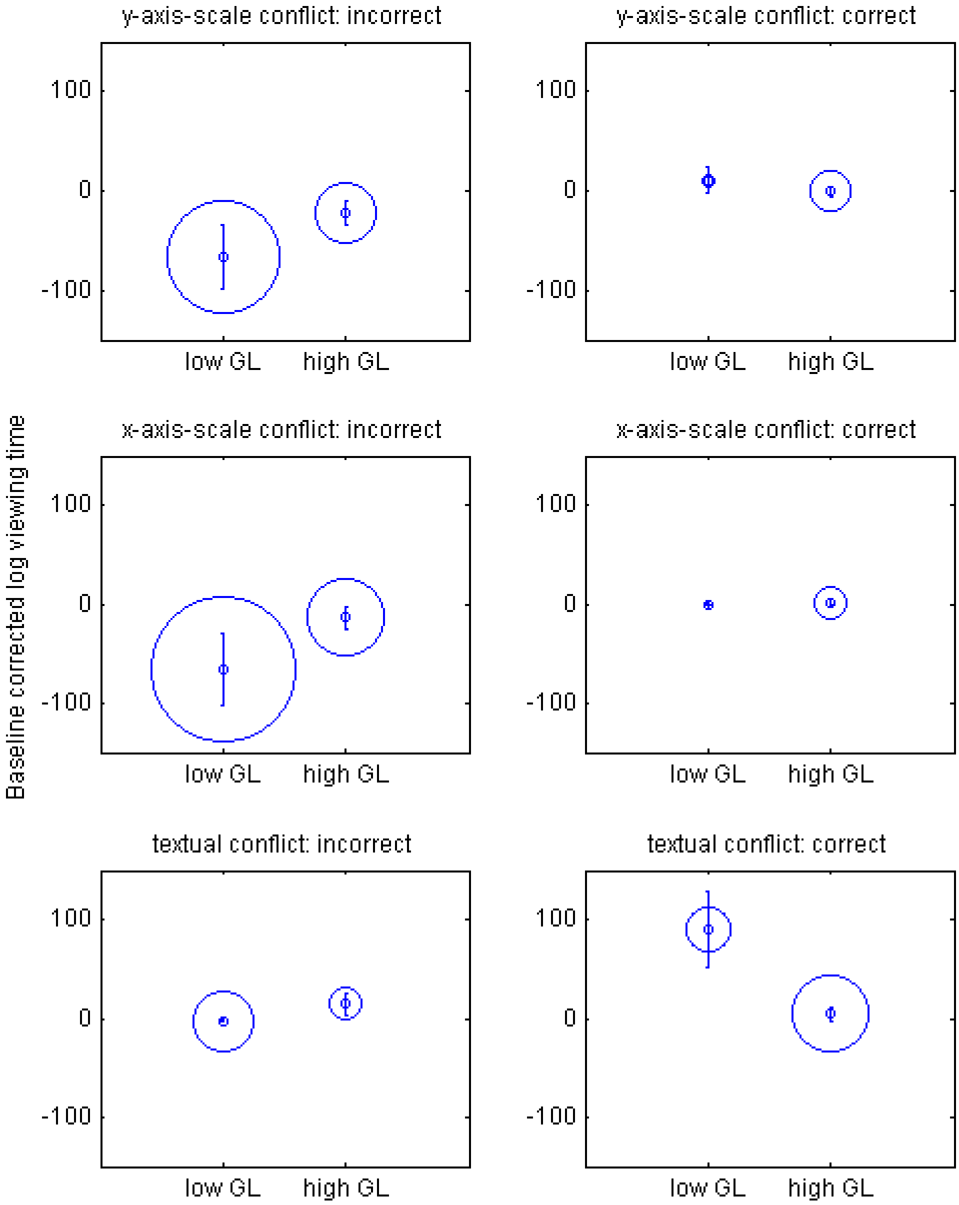
Figure 2. Average viewing times of less (low GL) and more (high GL) graph literate participants, by type of conflict (*y*-axis-scale, *x*-axis scale, or textual) and correctness of response (incorrect or correct). Viewing times for *y*-axis and *x*-axis conflict graphs correspond to *y*-axis and *x*-axis scales respectively, whereas viewing times for textual conflict graphs correspond to titles and *y*-axis labels. Inner circles denote average log-transformed viewing times. Bars denote +/-1 standard error. Areas of outer circles are proportional to the percentage of incorrect and correct responses for each graph literacy group and type of conflict.

Figure 1



y-axis-scale conflict: incorrect
y-axis-scale conflict: correct
textual conflict: incorrect
textual conflict: correct
Log viewing time
Log viewing time
1.5
1
0.5
0
-0.5
low GL
high GL

Figure 2


y-axis-scale conflict: incorrect
y-axis-scale conflict: correct
x-axis-scale conflict: incorrect
x-axis-scale conflict: correct
textual conflict: incorrect
textual conflict: correct
Baseline corrected log viewing time
100
0
-100
low GL
high GL

Appendix A

Graphs used in Experiments 1 and 2. *Note: Graphs G1, G2, G11, and G12 with conflicts were used in Experiment 1. All graphs were used in Experiment 2. In Experiment 1, response options for graphs G1 and G2 did not include "I can't say." Original text was in German.*

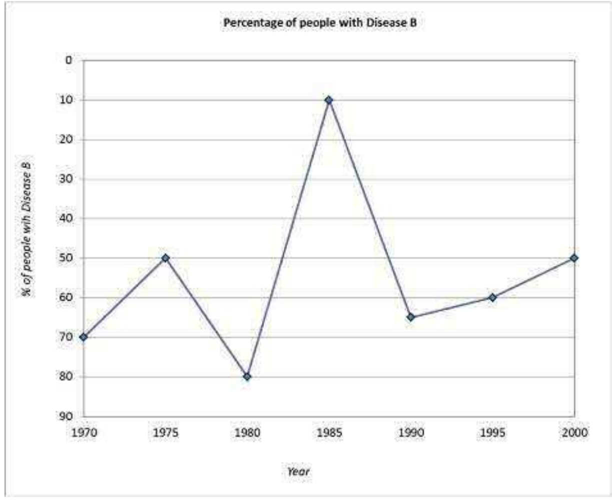
Graphs with conflict

Graphs without conflict

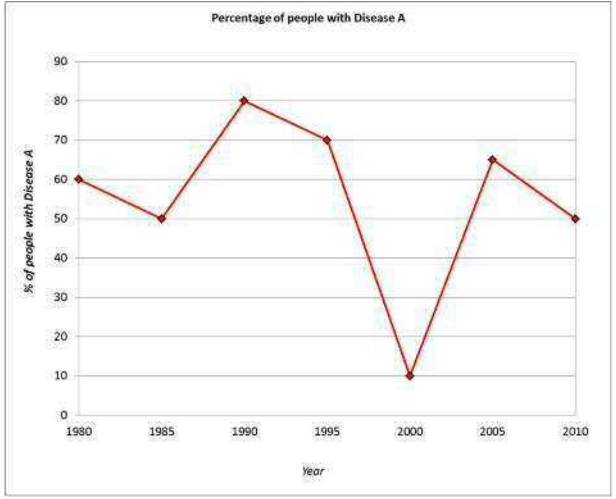
y-axis-scale conflict

G1. Inverted scale, *x* axis below

In which year was the percentage of people with Disease B highest?
(1) 1975 (2) 1980 (3) 1985 (4) 2000 (5) I can't say

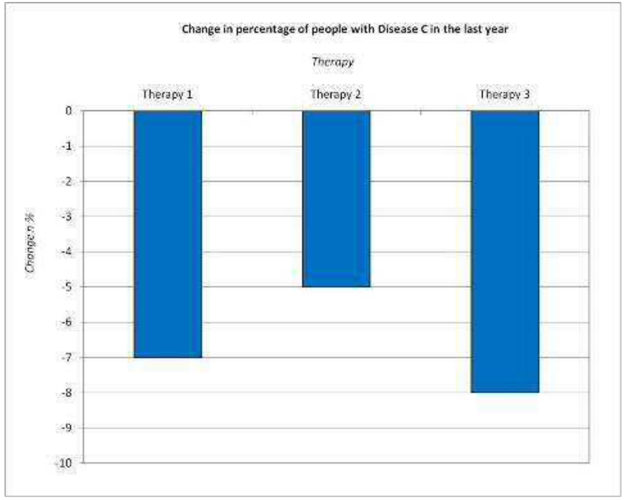


In which year was the percentage of people with Disease A highest?
(1) 1985 (2) 1990 (3) 2000 (4) 2005 (5) I can't say

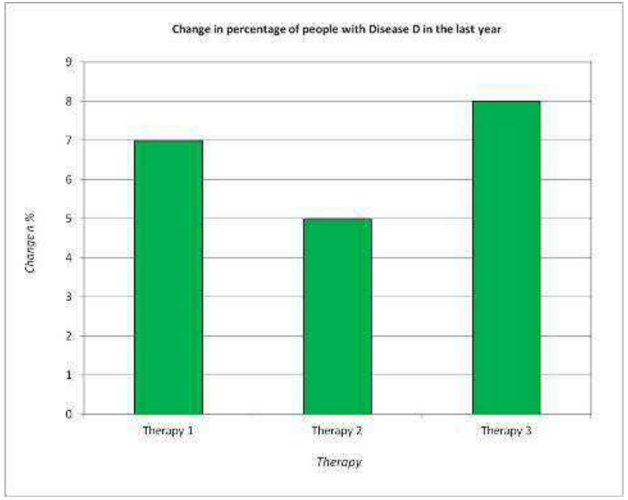


G2. Scale with negative values

For which type of therapy is the change in the percentage of patients with Disease C lowest? (1) Therapy 1 (2) Therapy 2 (3) Therapy 3 (4) It is the same for all therapies (5) I can't say



For which type of therapy is the change in the percentage of patients with Disease D lowest? (1) Therapy 1 (2) Therapy 2 (3) Therapy 3 (4) It is the same for all therapies (5) I can't say



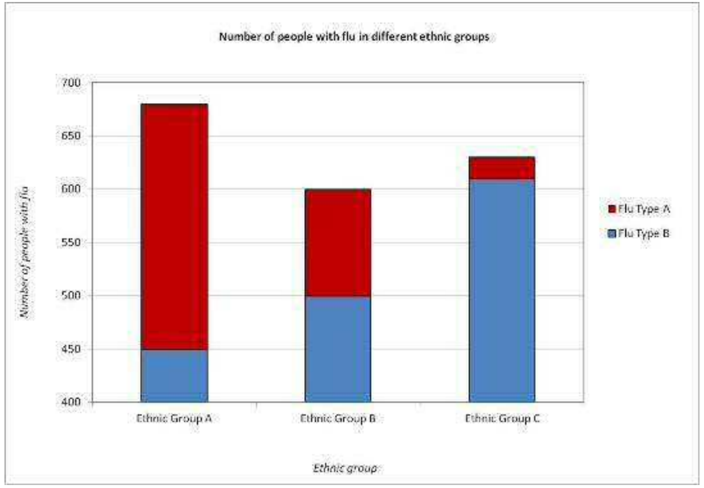
Graphs with conflict

Graphs without conflict

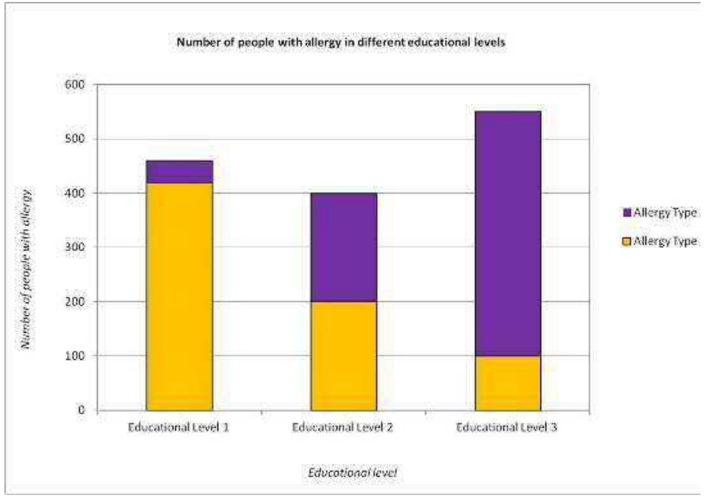
y-axis-scale conflict (continued)

G3. Excised scale

For which ethnic group was the proportion of people with flu Type A larger than the proportion of people with flu Type B? (1) Ethnic Group A (2) Ethnic Group B (3) Ethnic Group C (4) it was never larger (5) I can't say

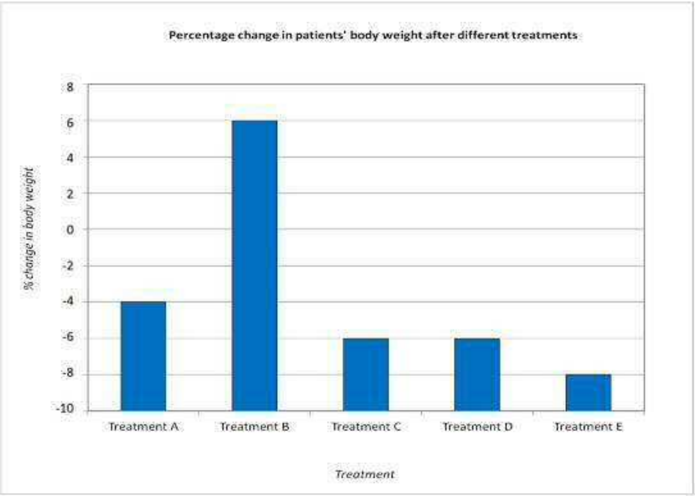


For which educational level was the proportion of people with allergy Type A larger than the proportion of people with allergy Type B? (1) Educational Level 1 (2) Educational Level 2 (3) Educational Level 3 (4) it was never larger (5) I can't say

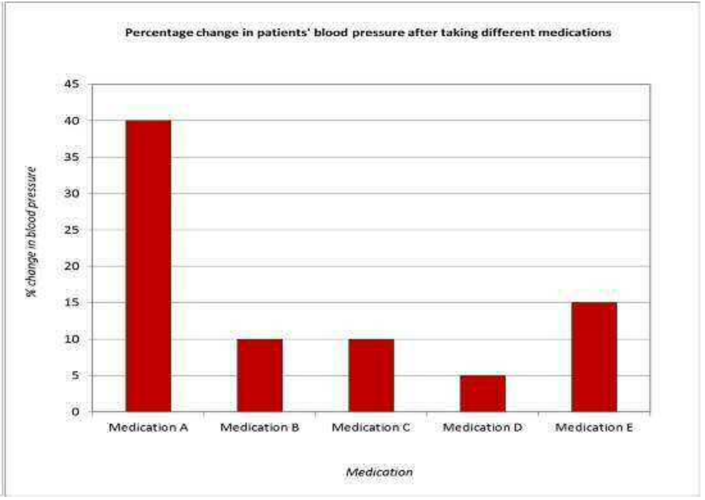


G4. Scale not indicating zero baseline

Which treatment resulted in the smallest change in patients' body weight?
(1) Treatment A (2) Treatment B (3) Treatment C (4) Treatment E (5) I can't say



Which medication resulted in the smallest change in patients' blood pressure?
(1) Medication A (2) Medication B (3) Medication C (4) Medication D (5) I can't say



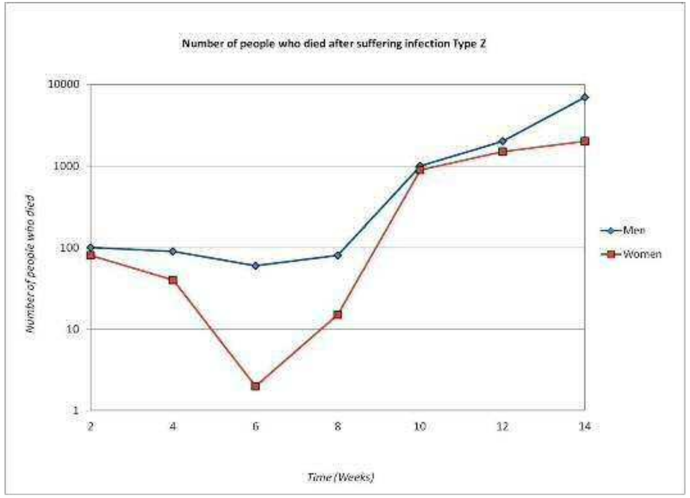
Graphs with conflict

Graphs without conflict

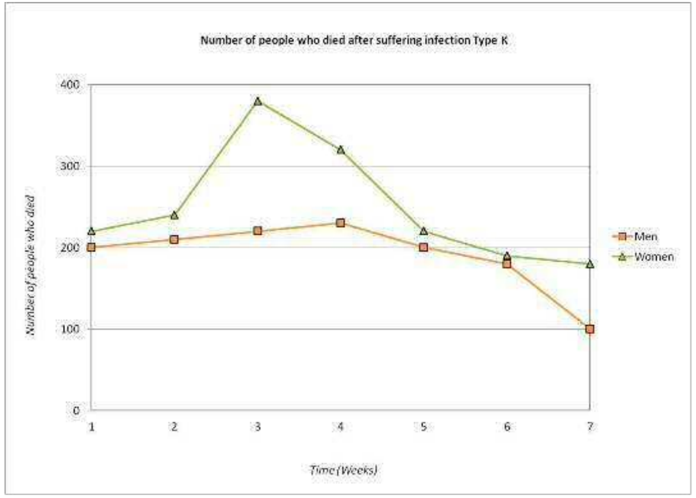
y-axis-scale conflict (continued)

G5. Logarithmic scale

When was there a larger difference between the number of men and women dying after suffering infection Type Z?
(1) Week 2 (2) Week 6 (3) Week 10 (4) Week 14 (5) I can't say

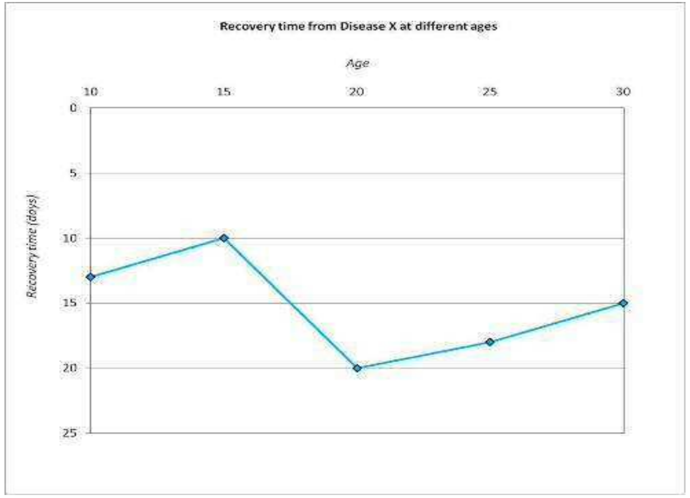


When was there a larger difference between the number of men and women dying after suffering infection Type K?
(1) Week 1 (2) Week 3 (3) Week 5 (4) Week 7 (5) I can't say

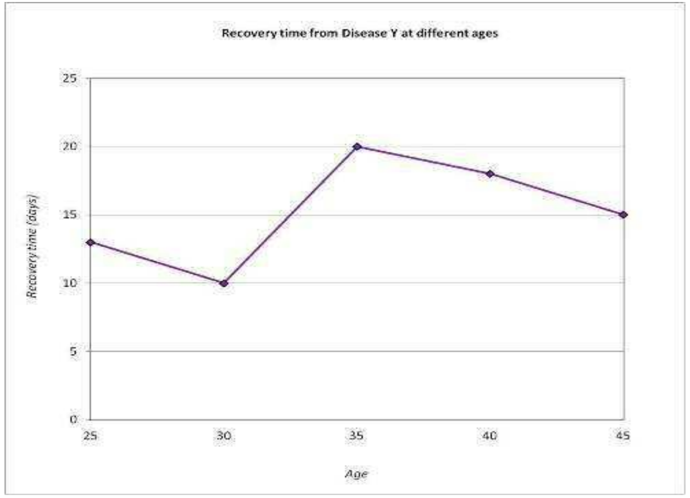


G6. Inverted scale, *x* axis above

At which age was the recovery time from Disease X lowest?
(1) At 15 (2) At 20 (3) At 25 (4) At 30 (5) I can't say



At which age was the recovery time from Disease Y lowest?
(1) At 25 (2) At 30 (3) At 35 (4) At 45 (5) I can't say

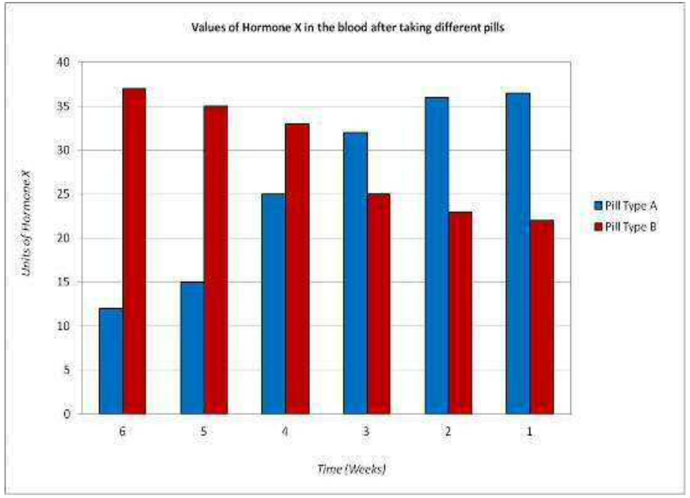


Graphs with conflict | **Graphs without conflict**

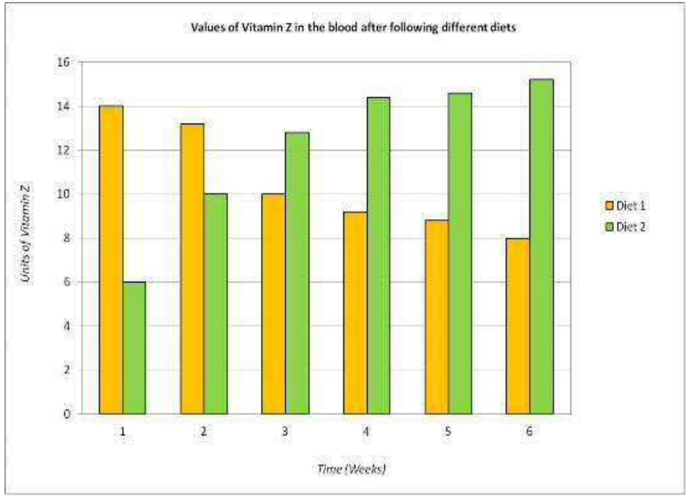
x-axis-scale conflict

G7. Inverted scale (bar)

Which pill resulted in an increase in the values of Hormone X in the blood over time?
(1) Pill Type A (2) Pill Type B (3) Both pills (4) Neither of the pills (5) I can't say

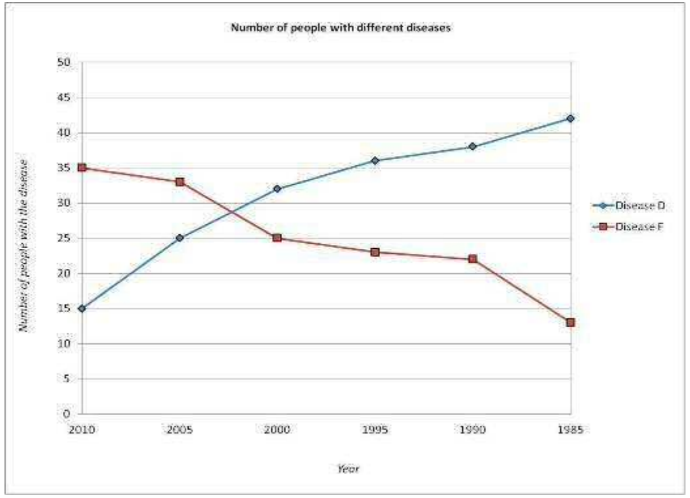


Which diet resulted in an increase in the values of Vitamin Z in the blood over time?
(1) Diet 1 (2) Diet 2 (3) Both diets (4) Neither of the diets (5) I can't say

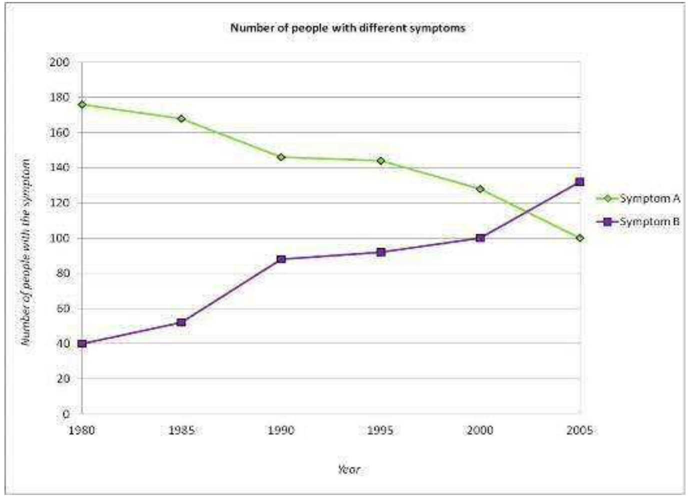


G8. Inverted scale (line)

For which disease did the number of affected people increase over time?
(1) Disease D (2) Disease F (3) For both of them (4) For neither of them (5) I can't say



For which symptom did the number of affected people increase over time?
(1) Symptom A (2) Symptom B (3) For both of them (4) For neither of them (5) I can't say



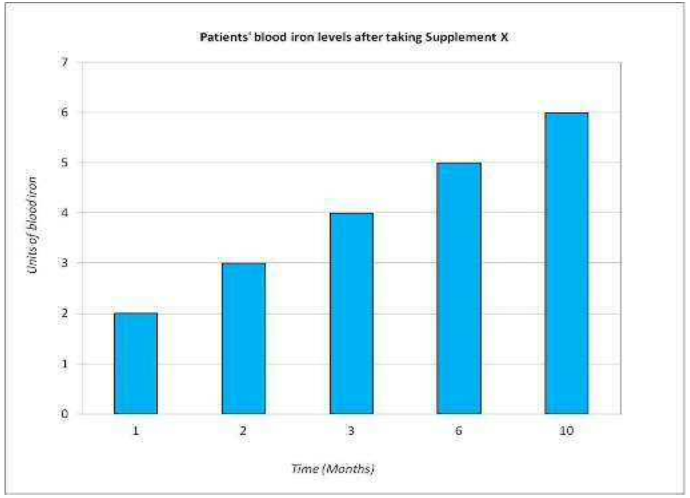
Graphs with conflict

Graphs without conflict

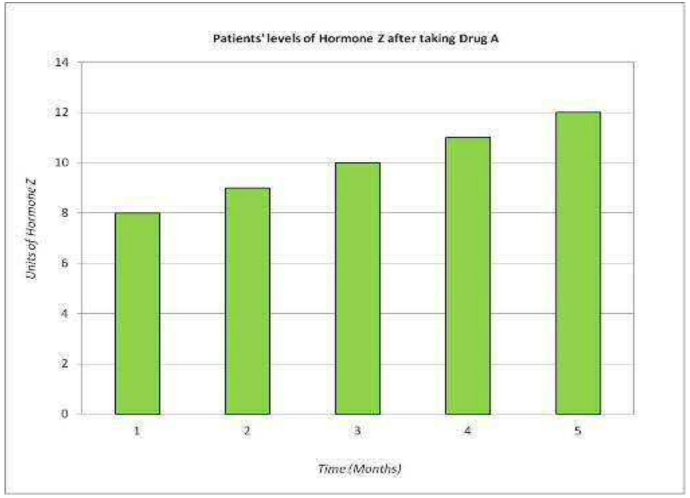
x-axis-scale conflict (continued)

G9. Values not placed at proportional distances (bar)

When did patients' blood iron levels start to increase more slowly? (1) After Month 2 (2) After Month 3 (3) After Month 6 (4) They increased equally quickly across all months (5) I can't say

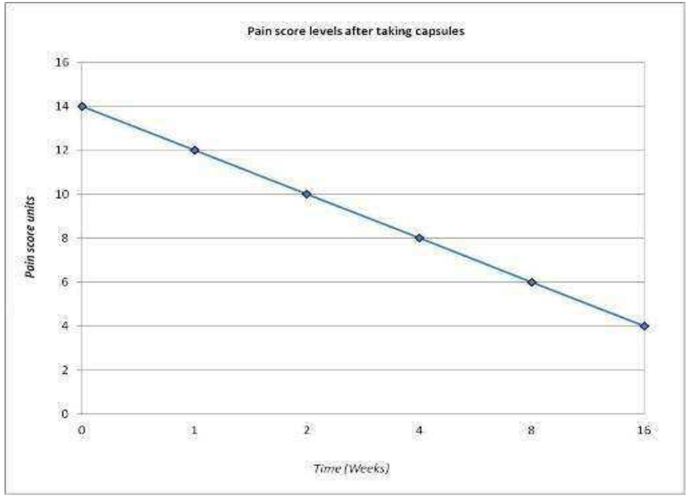


When did patients' levels of hormone Z start to increase more slowly? (1) After Month 2 (2) After Month 3 (3) After Month 4 (4) They increased equally quickly across all months (5) I can't say

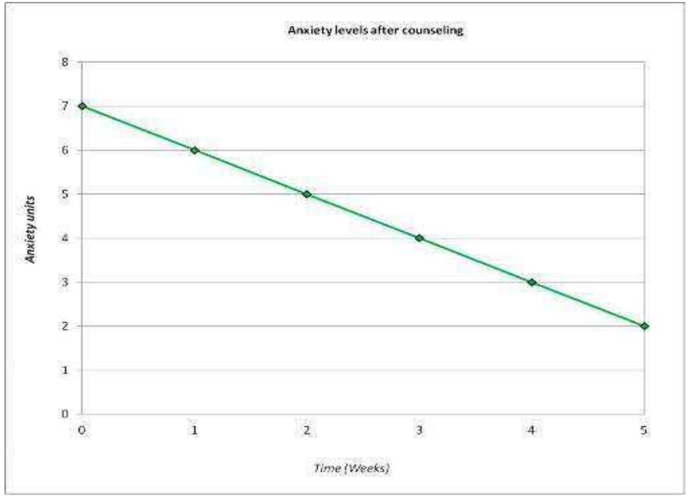


G10. Values not placed at proportional distances (line)

When did pain scores start decreasing more slowly? (1) After Week 1 (2) After Week 2 (3) After Week 4 (4) They decreased at the same pace across all weeks (5) I can't say



When did anxiety levels start decreasing more slowly? (1) After Week 1 (2) After Week 2 (3) After Week 4 (4) They decreased at the same pace across all weeks (5) I can't say



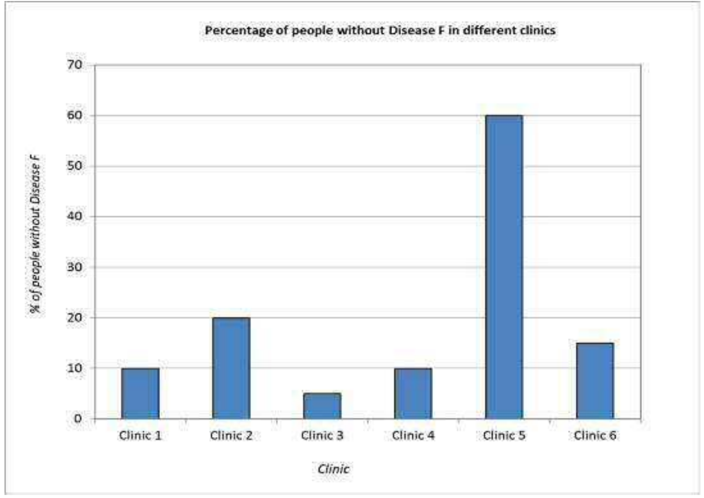
Graphs with conflict | **Graphs without conflict**

Textual conflict

G11. People without disease

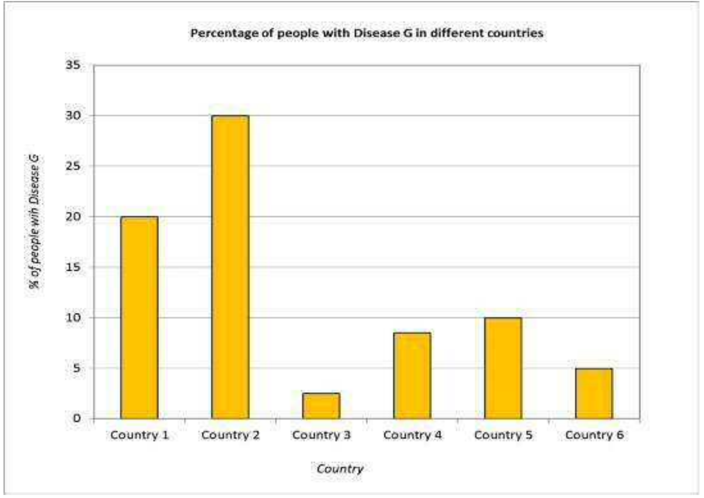
Graphs with conflict

In which clinic was the percentage of people with Disease F highest?
(1) Clinic 2 (2) Clinic 3 (3) Clinic 5 (4) Clinic 6 (5) I can't say



Graphs without conflict

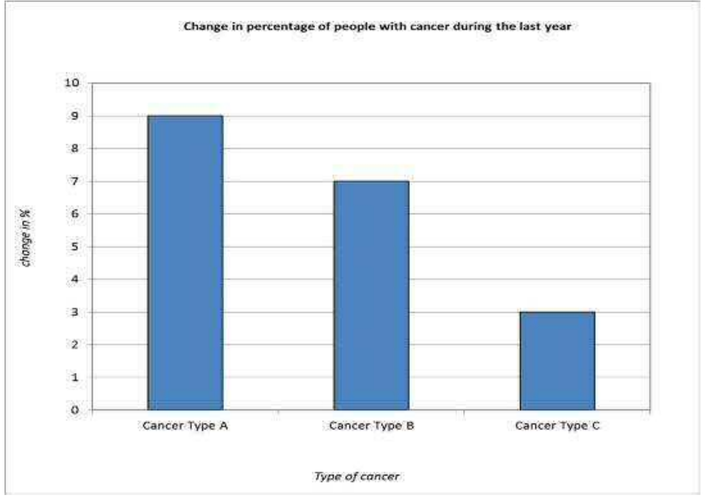
In which country was the percentage of people with Disease G highest?
(1) Country 1 (2) Country 2 (3) Country 3 (4) Country 5 (5) I can't say



G12. Change in the percentage of people with different types of cancer

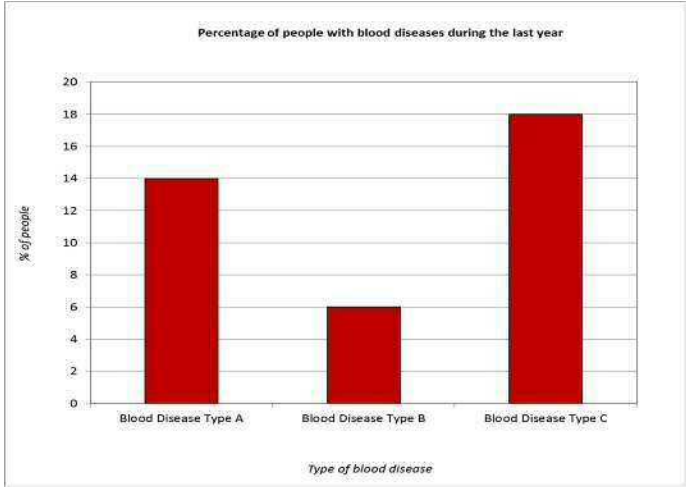
Graphs with conflict

Which type of cancer affected the smallest percentage of people during the last year?
(1) Cancer Type A (2) Cancer Type B (3) Cancer Type C (4) They are equal (5) I can't say



Graphs without conflict

Which type of blood disease affected the smallest percentage of people during the last year? (1) Blood disease Type A (2) Blood disease Type B (3) Blood disease Type C (4) They are equal (5) I can't say

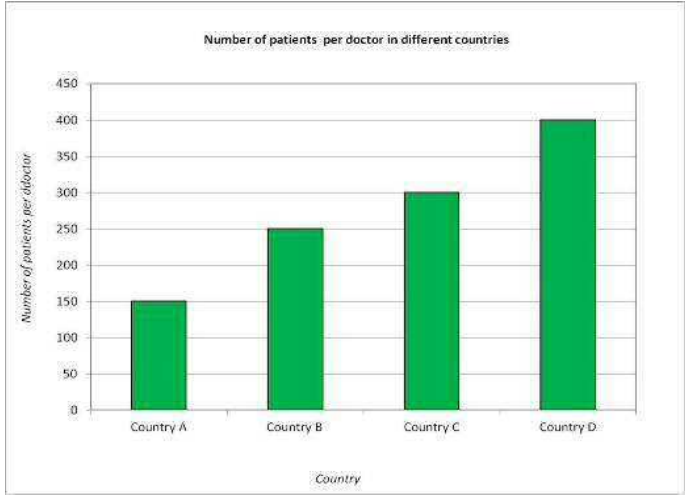


Graphs with conflict **Graphs without conflict**

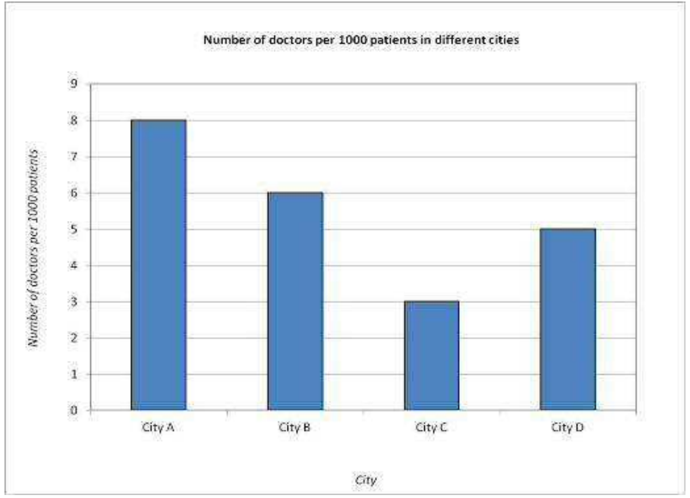
Textual conflict (continued)

G13. Number of patients per doctor

Which country has the highest number of doctors per patient?
(1) Country A (2) Country B (3) Country C (4) Country D (5) I can't say

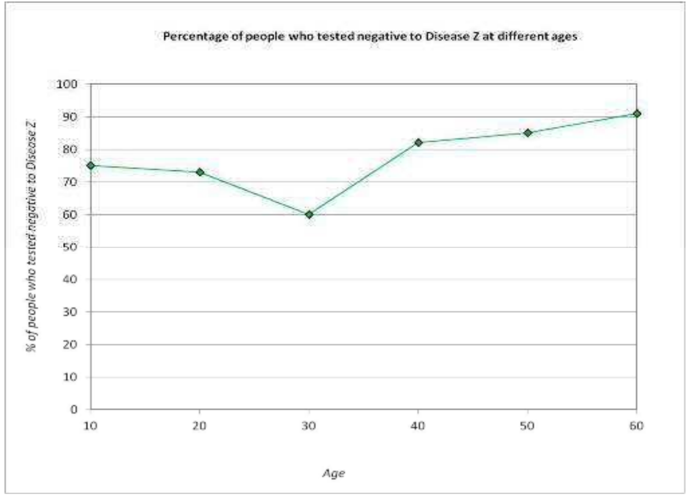


Which city has the highest number of doctors per patient?
(1) City A (2) City B (3) City C (4) City D (5) I can't say

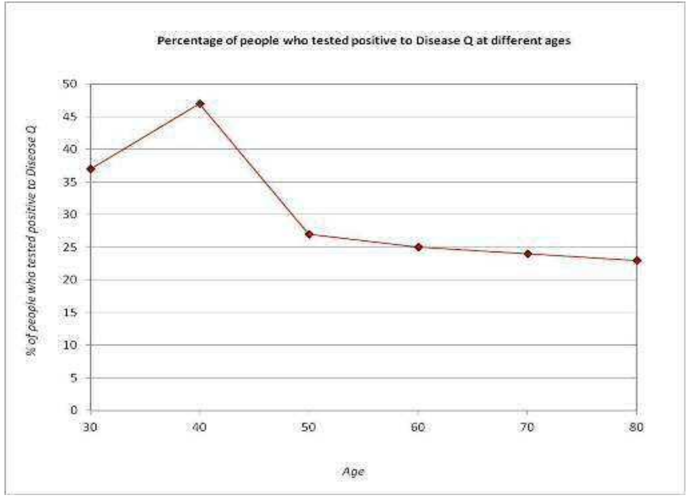


G14. People testing negative

At which age is the percentage of people diagnosed with Disease Z highest?
(1) At 10 (2) At 30 (3) At 50 (4) At 60 (5) I can't say



At which age is the percentage of people diagnosed with Disease Q highest?
(1) At 30 (2) At 40 (3) At 50 (4) At 80 (5) I can't say



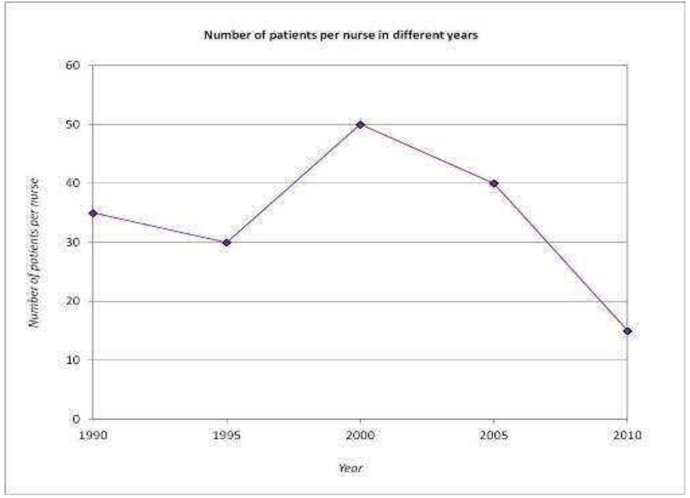
Graphs with conflict

Graphs without conflict

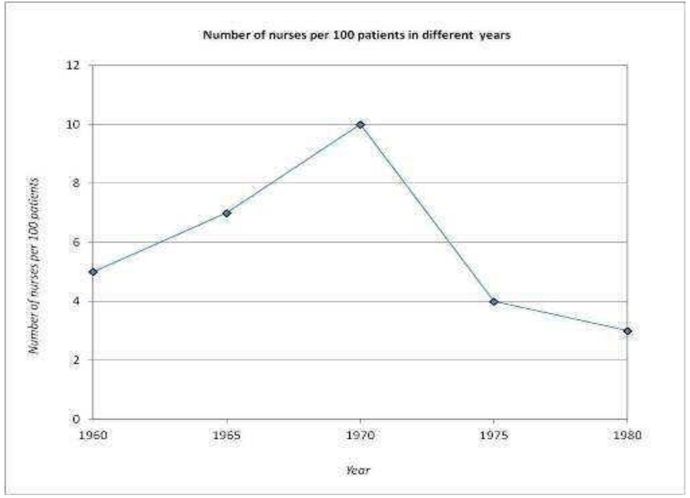
Textual conflict (continued)

G15. Number of patients per nurse

In which year was the number of nurses per patient lowest?
(1) 1995 (2) 2000 (3) 2005 (4) 2010 (5) I can't say

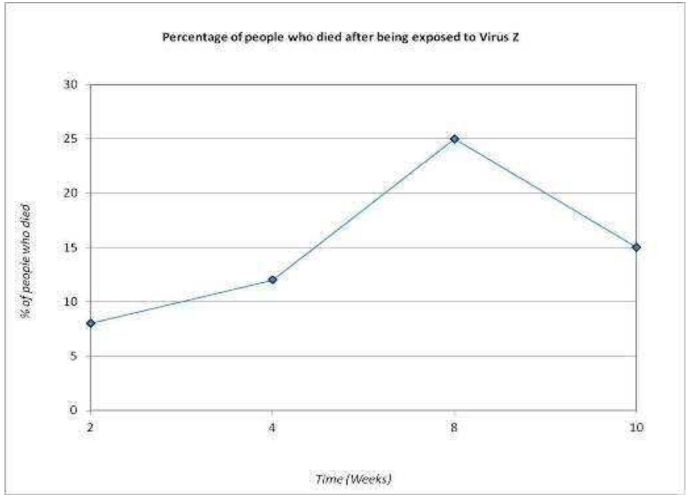


In which year was the number of nurses per patient lowest?
(1) 1960 (2) 1970 (3) 1975 (4) 1980 (5) I can't say

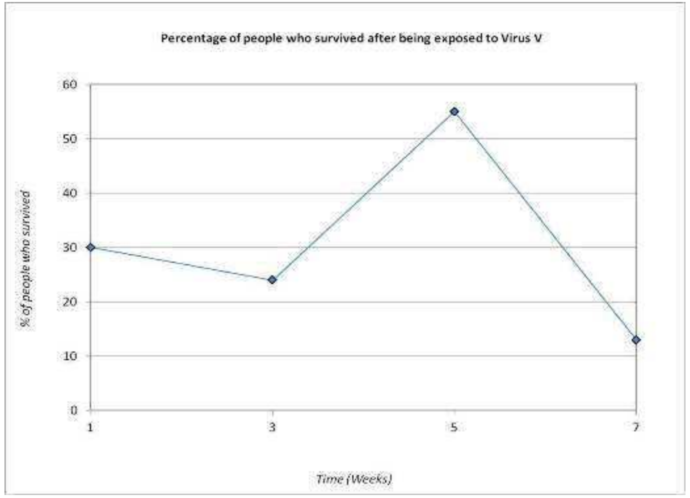


G16. People who died after virus exposure

When was the percentage of people who survived after being exposed to Virus Z lowest? (1) Week 2 (2) Week 4 (3) Week 8 (4) Week 10 (5) I can't say



When was the percentage of people who survived after being exposed to Virus V lowest? (1) Week 1 (2) Week 3 (3) Week 5 (4) Week 7 (5) I can't say



Appendix B

Sizes of the AOIs corresponding to the relevant conventional features, for graphs with conflicts. *Note: Values are shown in pixels. Sizes of AOIs for graphs without conflicts were identical to those for graphs with conflicts in the case of y-axis and x-axis scale conflicts, and were very similar in the case of graphs with textual conflicts.*

Item	AOI	Length	Height	Area
y-axis-scale conflict				
G1	*y*-axis scale	75	444	33300
G2	*y*-axis scale	75	444	33300
G3	*y*-axis scale	75	444	33300
G4	*y*-axis scale	75	444	33300
G5	*y*-axis scale	75	444	33300
G6	*y*-axis scale	75	444	33300
x-axis-scale conflict				
G7	*x*-axis scale	603	77	46431
G8	*x*-axis scale	625	72	45000
G9	*x*-axis scale	656	75	49200
G10	*x*-axis scale	656	75	49200
Textual conflict				
G11	*y*-axis label	74	267	68174
	title	544	89	
G12	*y*-axis label	70	178	60276
	title	556	86	
G13	*y*-axis label	74	234	57993
	title	447	91	
G14	*y*-axis label	73	420	86012
	title	629	88	
G15	*y*-axis label	74	317	65041
	title	501	83	
G16	*y*-axis label	76	234	76932
	title	558	106	

Appendix C

Further analyses: Transitions between global areas.

To broaden our exploration of the patterns of eye fixations in the current study, we further defined a set of areas of interest (AOIs) for all graphs that corresponded to the global elements of bar charts and line plots outlined in previous research (see, e.g., Carpenter & Shah, 1998; Kosslyn, 2006). Specifically, we divided the graphs into four global parts: the pattern, the *x* axis, the *y* axis, and the title. For this division, the *x* axis and *y* axis included the respective *x*-axis and *y*-axis values and labels. Following Carpenter and Shah (1998; see also Huestegge & Philipp, 2011), we then computed the number of transitions between these global areas. A transition was counted each time the participant broke a sequence of consecutive fixations on a given AOI to fixate on a different AOI. The question was also included as an AOI. Figure C1 shows the types of transitions made between the different global areas and how often each type occurred across graphs. In Experiment 1, the mean number of transitions across graphs was 18.3 (*SE* = 0.9), while in Experiment 2 it was 20.6 (*SE* = 1.1) for graphs with conflicts and 18.5 (*SE* = 0.8), for graphs without conflicts. In all cases, the most frequent types of transition across graphs were those between the pattern and the question, and between the pattern and the *x* axis (Figure B1). These results are in line with Carpenter and Shah's integrative model (1998), which predicts that a large proportion of transitions occur between the pattern and regions used to determine referents (e.g., *x* and *y* axes), as consequence of viewers' need to integrate information across these regions. No reliable differences were observed between participants with low and high graph literacy in the frequencies of the different types of transitions.

Figure C1. (a) The proportions of transitions made by participants between different global areas in Experiment 1; (b) The proportions of transitions made by participants between different global areas for graphs with conflicts in Experiment 2; (c) The proportions of transitions made by participants between different global areas for graphs without conflicts in Experiment 2.

Figure C1a

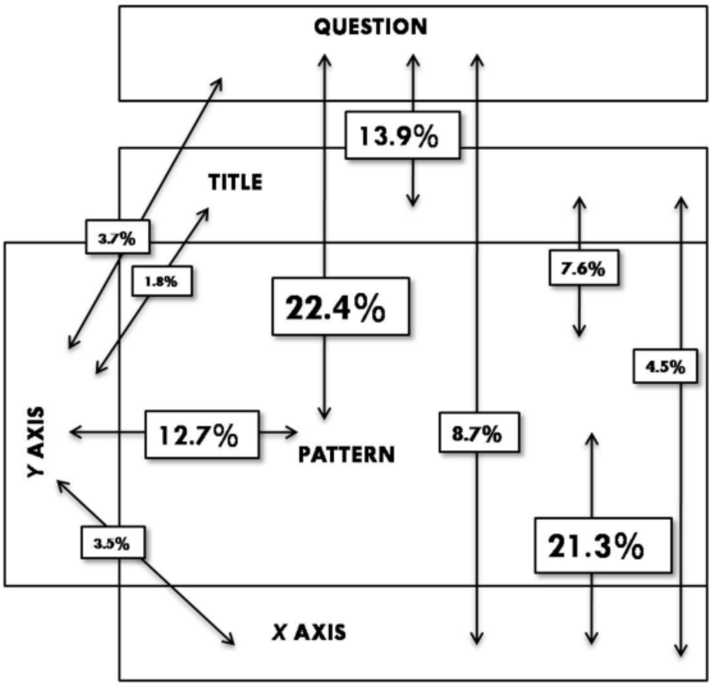


Figure C1b

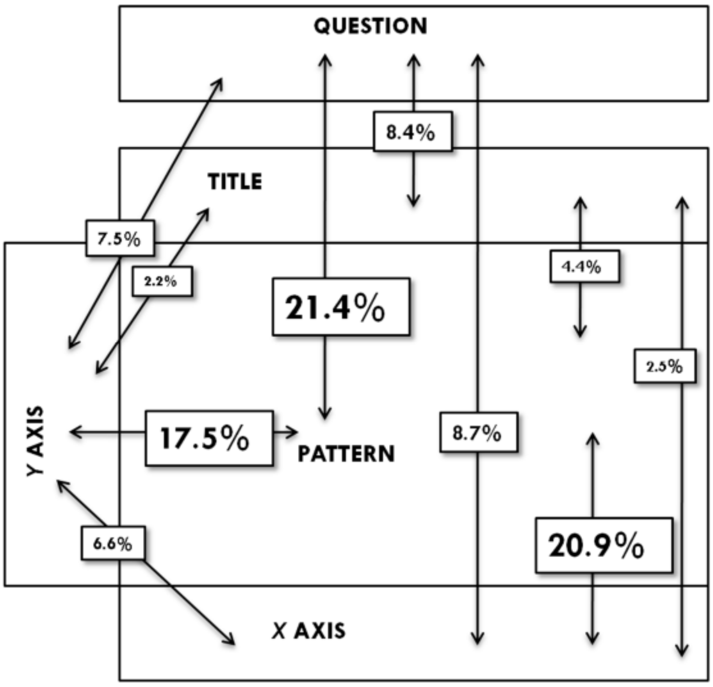
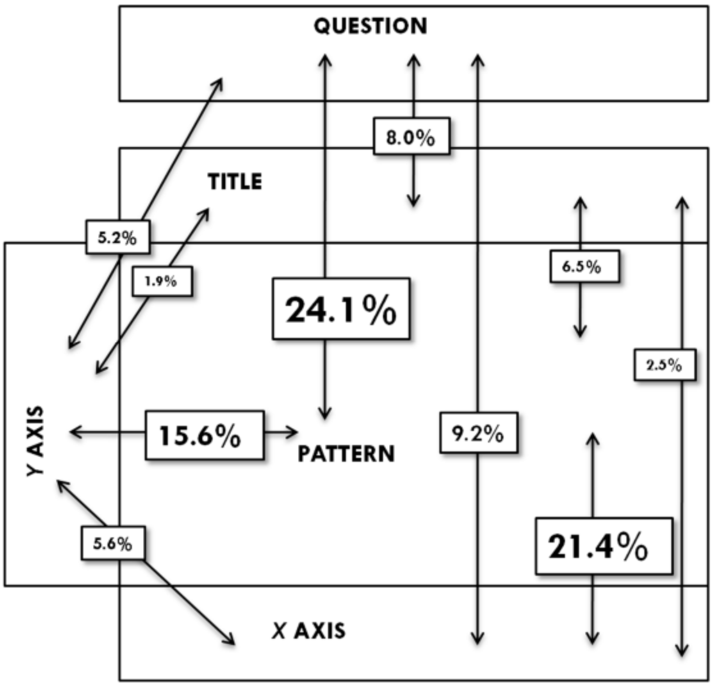

QUESTION
8.4%
TITLE
7.5%
2.2%
21.4%
4.4%
2.5%
8.7%
Y AXIS
17.5%
PATTERN
6.6%
20.9%
X AXIS

Figure C1c



QUESTION
TITLE
Y AXIS
PATTERN
X AXIS
8.0%
5.2%
1.9%
24.1%
6.5%
2.5%
15.6%
9.2%
5.6%
21.4%

Appendix D

Items measuring knowledge that graphs can be misleading. *Note: The response options provided were "Yes" and "No." Yes responses were coded with 1 and No responses with 0, for a total possible score of 6.*

Thinking about graphs that you might have seen in different contexts, such as graphs presenting data for different financial, nutritional, or political options and trends …

Do you think they are sometimes designed in a way that…

1. Makes some options look better or worse than they really are (e.g., by making differences in the data presented look larger or smaller)?
2. Directs attention to a particular option or aspects of that option (e.g., by directing attention to specific values in the data)?
3. Makes trends look more positive or negative than they really are (e.g., by distorting or misrepresenting the trends in the data)?

Thinking about graphs that you might have seen presenting medical information, such as data for different treatments and screenings (e.g., results of medical trials and pharmaceutical advertisements)…

Do you think they are sometimes designed in a way that…

1. Makes some options look better or worse than they really are (e.g., by making differences in the data presented look larger or smaller)?
2. Directs attention to a particular option or aspects of that option (e.g., by directing attention to specific values in the data)?
3. Makes trends look more positive or negative than they really are (e.g., by distorting or misrepresenting the trends in the data)?