

Effects of spatial compatibility on integration processes in graph comprehension

Lynn Huestegge · Andrea M. Philipp

Published online: 16 June 2011
© Psychonomic Society, Inc. 2011

Abstract A precondition for efficiently understanding and memorizing graphs is the integration of all relevant graph elements and their meaning. In the present study, we analyzed integration processes by manipulating the spatial compatibility between elements in the data region and the legend. In [Experiment 1](#), participants judged whether bar graphs depicting either statistical main effects or interactions correspond to previously presented statements. In [Experiments 2 and 3](#), the same was tested with line graphs of varying complexity. In [Experiment 4](#), participants memorized line graphs for a subsequent validation task. Throughout the experiments, eye movements were recorded. The results indicated that data–legend compatibility reduced the time needed to understand graphs, as well as the time needed to retrieve relevant graph information from memory. These advantages went hand in hand with a decrease of gaze transitions between the data region and the legend, indicating that data–legend compatibility decreases the difficulty of integration processes.

Keywords Compatibility · Graph comprehension · Display design · Eye movements · Saccades

In all types of visual media, verbal information is often complemented by graphs, allowing communication of complex information within a small space (Larkin &

Simon, 1987). Today, computer software supports a vast array of possibilities for the visualization of data, including bar and line graphs (Harris, 1999). Graph design guidelines were often intuitively derived by postulating plausible principles—for example, maximizing the “data–ink ratio” (Tufte, 1983). However, in recent years, several design aspects have been empirically evaluated (e.g., Carpenter & Shah, 1998; Carswell, Frankenberger, & Bernhard, 1991; Fischer, 2000; Shah & Carpenter, 1995; Siegrist, 1996; Spence, 1990; Zacks, Levy, Tversky, & Schiano, 1998), and the present study was designed to focus on the issue of spatial compatibility between graph elements.

Compatibility has been subject to a long research tradition (e.g., Proctor & Vu, 2006), and it is widely known that compatibility enhances performance by decreasing the amount of errors and/or response times (RTs; see Hommel & Prinz, 1997; Kornblum, 1992). In the broader sense, compatibility comprises any matching relation between elements, ranging from physical features of stimuli and responses to more abstract concepts, such as purposes or expectations. When we apply this broader sense of compatibility to graph comprehension, the purpose for which it is to be used determines which display layout represents the most compatible choice (Sparrow, 1989; Véssey, 1991, 1994; Washburne, 1927; Wickens & Andre, 1990; Wickens & Carswell, 1995). Consequently, tables are used for communicating exact data (Meyer, 2000), line graphs for trends (Zacks & Tversky, 1999), and bar graphs for identifying maxima (Meyer, Shinar, & Leiser, 1997) or contrasts (Zacks & Tversky, 1999). Other studies have focused on the compatibility of graphs with expectations based on previous knowledge (see Pinker, 1990), showing that both familiar graph layouts and data patterns reflecting expectations based on real-world knowledge yielded faster

L. Huestegge (✉) · A. M. Philipp
Institute of Psychology, RWTH Aachen University,
Jägerstraße 17–19,
52056 Aachen, Germany
e-mail: lynn.huestegge@psych.rwth-aachen.de

and/or more accurate graph comprehension (Fischer, Dewulf, & Hill, 2005; Gattis & Holyoak, 1996; Shah, 1995).

According to a narrower sense, compatibility refers to the relation of the physical features of stimuli and responses (S–R compatibility) or of features of different elements in a stimulus display (S–S compatibility) (Fitts & Simon, 1952; Proctor & Vu, 2006). Unfortunately, compatibility in the narrower sense has been largely neglected in the study of graph comprehension. One exception is a study by Feeny, Holo, Liversedge, Findlay, and Metcalf (2000), in which the spatial compatibility of the order of words in a sentence and the location of corresponding graph elements was manipulated (see also Fischer, 2000). Recently, Renshaw, Finlay, Tyfa, and Ward (2004) reported an eye-tracking study in which they compared optimal versus suboptimal graph layouts. While the former included direct labeling of lines in line graphs, the latter involved spatially incompatible legend entries but, at the same time, lower overall contrast, additional irrelevant 3-D elements, and so forth. As a result, they found that eye movement patterns significantly differed between processing of the two types of graphs, demonstrating the usefulness of eye movement measurement in the evaluation of graph design. However, this study was not designed to selectively assess the contribution of incompatibility to the variance in eye movements. In the present study, we specifically focus on the influence of spatial compatibility versus incompatibility on graph comprehension.

Various theoretical accounts of graph comprehension (e.g., via task analyses) have been proposed (Cleveland & McGill, 1984, 1986, 1987; Gillan & Lewis, 1994; Lohse, 1991, 1993; Simkin & Hastie, 1987). An influential model of graph comprehension comprising three phases was introduced by Carpenter and Shah (1998). An initial pattern recognition phase is devoted to encoding, an interpretation phase involves retrieving qualitative and quantitative meanings, and an integration process relates meanings to semantic referents defined by labels, legends, or titles. These phases are considered to be integrative parts of cycles of recognition and interpretation, an assumption that was empirically supported by frequent gaze transitions between graph elements.

Unfortunately, integration processes in graph comprehension had, for a long time, “received little empirical attention” (Carpenter & Shah, 1998, p. 76). Earlier studies mainly demonstrated the importance of spatial design characteristics for integration processes within different task requirements (Carswell & Wickens, 1990; Gillan, 1995; Lohse, 1991). However, in the past 12 years, influential studies have emerged that have been explicitly devoted to integration processes (e.g., Bryant & Tversky, 1999; Zacks & Tversky, 1999). For example, a recent eye movement study by Ratwani, Trafton, and Boehm-Davis (2008) presented evidence that integration can be subdivided into visual integration (using perceptual features to build visual clusters) and cognitive

integration (higher-level comparison of clusters). These integration processes were shown to be more demanding as visual graph complexity increased. Another line of research utilizing oculomotor analyses demonstrated that integration is partly determined by graph types and user characteristics (Peebles, 2008; Peebles & Cheng, 2001, 2003).

One specific aspect of the integration phase is that information from the data region and the legend (and/or the title) needs to be integrated. Thus, for effective integration processes, any difficulties with regard to legend or title should be minimized—for example, by avoiding legends through direct labeling of bars and lines (Gillan, Wickens, Hollands, & Carswell, 1998; Kosslyn, 1994; Schmid, 1983). However, in complex or dense graphs, direct labeling easily results in visual cluttering of textual elements and, thus, cannot be considered a panacea. One possible difficulty with respect to the legend is its spatial incompatibility with the lines or bars in the graph. In the present study, we investigated the effects of spatial incompatibility by manipulating the spatial relation of entries between data region and legend. This issue seems especially important since graph-generating software often does not easily allow one to change the sequence of legend entries. Furthermore, many research reports and even research guidelines still include spatially incompatible legend–data relations (see American Psychological Association (APA) 2006, pp. 180 and 192; APA, 2010, p. 156, where incompatible legends are presented as a reference).

Thus, the aim of the present study was to test whether spatial incompatibility between the data region and the legend hampers graph comprehension and whether this effect is found across different display types (bar graphs vs. line graphs). To further increase the scope of the research question, we systematically varied various forms of complexity to further specify the conditions under which compatibility effects may occur during graph comprehension. More specifically, complexity was manipulated in three independent ways. First, in each experiment, we manipulated *data pattern complexity* by asking participants to evaluate either depicted statistical main effects or interactions. We reasoned that interactions are more complex, because they involve the comparison of differences between data points, instead of just comparing overall means (Halford, Wilson, & Phillips, 1998). **Experiment 3** additionally addressed the effect of *visual complexity* by increasing the amount of data depicted within a graph, in comparison with **Experiment 2**. **Experiment 4** dealt with the *complexity of task requirements* by having participants judge graphs from memory. Throughout the experiments, eye movements were recorded to further specify integration processes by deciding whether the expected adverse effects of incompatibility were due to an increased frequency of gaze transitions between the data region and the legend.

Despite the fact that numerous previous display compatibility studies clearly provided evidence for a general advantage of compatible layouts, the role of compatibility in data–legend integration appears less clear. If we assume that our cognitive system is organized in a perfectly efficient manner, legend inspection should result only in memory codes consisting of the relevant marker–label bindings, because their spatial position is taskirrelevant and can be considered as surface information (similar to surface information in text comprehension; see Kintsch, Welsch, Schmalhofer, & Zimny, 1990). Furthermore, working memory demands related to legend information should be rather low in graphs with only two legend entries, even in the presence of additional memory load related to the statements. Thus, one would not necessarily predict that participants easily forget about the legend content, subsequently being forced to revisit the legend with their gaze. Consequently, this view would not inevitably imply adverse effects of data–legend incompatibility under all conditions. Therefore, the present research goal was to determine under which conditions data–legend compatibility would matter in graph comprehension.

To anticipate the results, we found strong evidence for an adverse effect of data–legend incompatibility on graph comprehension with respect to both RTs and number of gaze transitions whenever a certain level of cognitive complexity was crossed. Thus, we suppose that compatibility between data region and legend can play a crucial role for integration processes in graph comprehension.

Experiments 1 and 2

In [Experiment 1](#), we examined legend compatibility in bar graphs, a display type often recommended for depicting main effects and the comparison of contrasts (interactions). To enable the manipulation of compatibility, the legend was placed above the data region, representing a common option in spreadsheet software. In the compatible condition, right bars and right legend entries referred to the same data, whereas in the incompatible condition, legend entries were spatially reversed. In [Experiment 2](#), we used the same data material but created line graphs with legends presented on the right side. Despite the use of the same data material, we decided against the presentation of both experiments within a single experimental design. The reason for this was that any differences between line and bar graphs could not be attributed only to the general type of graph (bars vs. lines), but also to the specific layout of each graph type, which involved a number of arbitrary decisions regarding the design of the line markers, the width and spatial arrangement of the bars, and so forth.

Method

Participants Eighteen students (14 of them female and 4 male) with a mean age of 26 years (range: 20–48) participated in [Experiment 1](#). Sixteen different students (14 of them female and 2 male) with a mean age of 22 years (range: 20–25) participated in [Experiment 2](#). They had normal or corrected-to-normal vision. Their prior experience with graphs (here and in the following experiments) was about equal to that of an advanced BSc psychology student.

Stimuli Graphs (generated with Microsoft Excel) in [Experiment 1](#) depicted either main effects or ordinal interactions. Each consisted of four bars (two gray and two black) representing *screen viewing time* on the *y*-axis. The *x*-axis referred to the independent variable *age* and depicted children on the left and adults on the right (see [Fig. 1](#)). For each age group, separate bars (gray vs. black) represented the independent variable *screen type* (TV vs. computer) defined through the legend. The legend ($24^\circ \times 1^\circ$ of visual angle) was presented in one line above the data region ($24^\circ \times 13^\circ$). The spatial separation between legend and data region amounted to 2° .

We designed four basic graph figures. For graphs depicting main effects, we varied the overall difference between children and adults (large vs. small). For graphs depicting interactions, we varied whether the difference in bar height was more pronounced for the left than for the right bars. These four basic figures served as templates: For each figure, we varied the position of the bar colors (black left/gray right and vice versa), the position of each legend marker (black/gray square) in the legend (left vs. right), and the order of legend entries (“TV” left/“computer” right and vice versa), resulting in 32 figures with a compatible or incompatible data–legend relation. For each value on the *x*-axis, smaller bars were always placed on the left side to avoid the introduction of an additional variable, as compared with the line graphs in [Experiment 2](#).

Each trial consisted of the presentation of one out of four statements (black on gray background) extending over two text lines (30° horizontally) and a subsequently shown bar graph. Graphs depicting a main effect were preceded by one of two respective statements (e.g., “In general, people spent more time in front of the computer[TV] than in front of the TV[computer]”), whereas interaction graphs were preceded by interaction statements (e.g., “The difference in viewing times between children and adults is larger for the TV [computer] than for the computer[TV]”). Each of the 32 graphs was combined with a matching and a nonmatching statement, resulting in 64 experimental trials altogether.

Stimuli for [Experiment 2](#) were designed using the same data, which were converted into grayscale line graphs

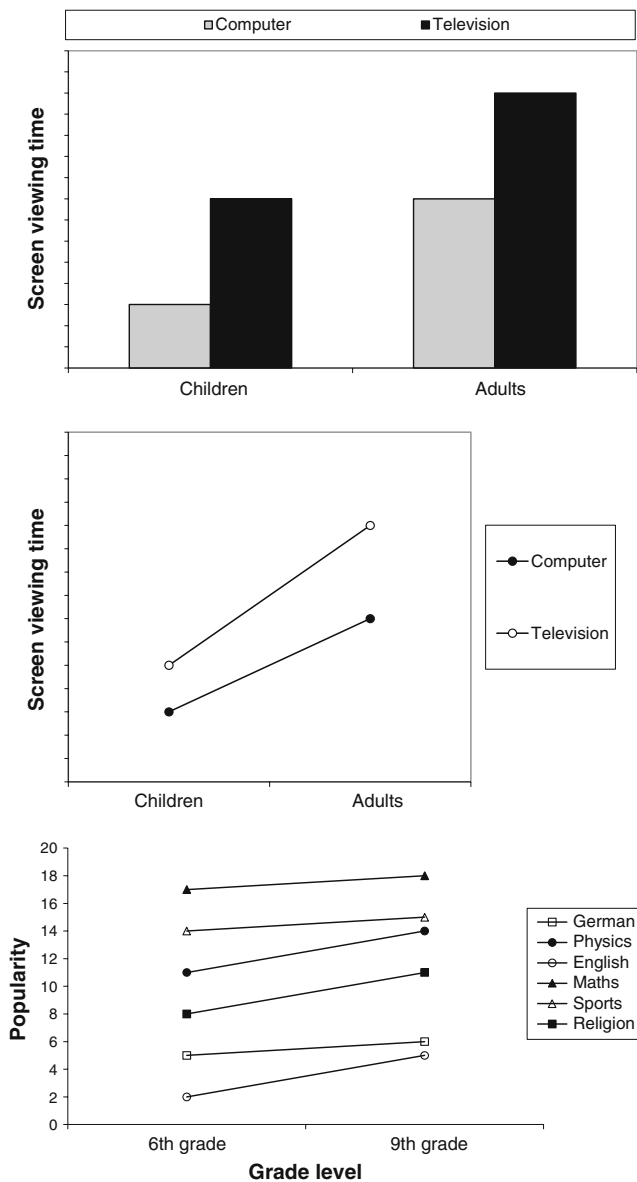


Fig. 1 Examples of graphs used in Experiments 1 (upper graph, compatible), 2 and 4 (middle graph, incompatible), and 3 (lower graph, incompatible). Original graphs were depicted in German. Due to software conversion issues, reproductions might slightly differ from the originals used in the study

consisting of two uncrossed black lines. Data point markers were either black or gray circles. The legend ($6^\circ \times 4^\circ$) was placed on the right side of the data region ($18^\circ \times 16^\circ$) and contained two lines, each consisting of a marker (black/gray circle) and a label (“TV”/“computer”; see Fig. 1). The spatial separation between legend and data region amounted to 4° .

Apparatus, task, and procedure Statements and figures were presented centrally on a 21-in. CRT monitor (100 Hz, $1,240 \times 1,068$ pixels), with a viewing distance of 67 cm. We utilized a 500-Hz EyeLink II head-mounted

video-based eyetracker (SR Research, Canada) with a chinrest. By using infrared reflection, the eyetracker records the position of the pupil, resulting in an average spatial accuracy below 0.5° . Nine-point calibration routines were conducted each ten trials throughout the experiment.

The experiment was run in a single session of 20 min and consisted of a visual instruction followed by 4 practice trials (to accommodate participants to the task) and 64 randomized experimental trials. Each trial started with the presentation of the statement. Participants ended viewing time by pressing the space bar of a German QWERTZ keyboard (left index finger). This was followed by a black fixation cross (height and width, 0.5°) presented for 250 ms, placed in between the data pattern and the legend of the subsequent figure. Then the graph was presented, and participants decided via keypress (right index and middle fingers on mouse buttons) as quickly and accurately as possible whether it corresponded to the previous statement. Meaning of the response keys (match vs. mismatch) was counter balanced across participants. A blank screen was presented for 600 ms after each response had been given. No feedback was provided.

Design Compatibility (legend spatially compatible vs. incompatible with data region) and the depicted effect type (main effect vs. interaction depicted in the graph) were within-subjects independent variables. Manual RTs (for correct trials) and accuracy (percentage correct) were dependent variables (performance measures). For the sake of readability, we restricted gaze analyses to the most informative parameter—namely, the frequency of gaze transitions between the data region and legend. Gazes were defined as the sum of fixations on a graph region until the eyes left the region. A gaze transition was registered whenever a saccade started in one of the two regions (plus an extra margin of 0.5° of visual angle for each region, to compensate for potential spatial inaccuracy of the gaze-tracking data) and ended in the other. We did not report fixation frequencies or durations for legends, since these parameters were highly correlated with gaze transitions and yielded the same overall pattern of results throughout all the experiments. We mainly conducted two-way repeated measurement ANOVAs ($\alpha = .05$ throughout), sometimes complemented by post hoc *t*-tests (Bonferroni corrected) to further qualify interactions.

Results

Experiment 1 (bar graphs) Trials with exceedingly long statement- or graph-viewing times ($+3 SD$) were excluded (corresponding to seven trials altogether). RTs were greater for incompatible legends than for compatible legends, $F(1, 17) = 13.63$, $p = .002$, $\eta_p^2 = .45$, and for the judgment

of depicted interactions, as compared with depicted main effects, $F(1, 17) = 14.38, p = .001, \eta_p^2 = .46$ (see Fig. 2). There was a significant interaction of compatibility and the depicted effect type, $F(1, 17) = 5.79, p = .028, \eta_p^2 = .25$. Posthoc t -tests revealed that there was no significant RT effect of compatibility for the judgment of main effects (compatible, $M = 3.33$ s, $SE = 0.36$; incompatible, $M = 3.42$ s, $SE = 0.36$) $t < 1$, but there was for interactions (compatible, $M = 3.85$ s, $SE = 0.38$; incompatible, $M = 4.83$ s, $SE = 0.57$), $t(17) = 3.21, p = .005$.

Accuracy, computed as the percentage of correct responses for each participant, amounted to 83.90% ($SE = 3.10$) for compatible graphs, as compared with 76.40% ($SE = 4.70$) for incompatible graphs, yielding a significant effect of compatibility, $F(1, 17) = 5.15, p = .037, \eta_p^2 = .23$. Accuracy was greater for the judgment of depicted main effects (87.1%), as compared with interactions (76.4%), $F(1, 17) = 12.94, p = .002, \eta_p^2 = .43$. There was no significant interaction of depicted effect type and compatibility, $F < 1$.

Eye movements were analyzed regarding the amount of gaze transitions from the data region to the legend and vice versa. The mean number of gaze transitions between data region and legend was significantly higher during the inspection of incompatible graphs ($M = 2.68$ transitions, $SE = 0.41$), as compared with compatible graphs ($M = 2.39$ transitions, $SE = 0.39$), $F(1, 17) = 6.05, p = .025, \eta_p^2 = .26$. There was no main effect of the depicted effect type, and no significant interaction between depicted effect type and compatibility (both F s < 1). A further analysis revealed that, overall, gaze transitions from legend to data region occurred more often ($M = 1.37, SE = 0.20$) than gaze transitions from data region to legend ($M = 1.17, SE = 0.20$), $t(17) = 3.18, p = .005$. The mean percentage of trials on which the legend was fixated before the data region across participants amounted to 89.4% ($SE = 3.38$).

To test whether RTs and number of gaze transitions were directly related, we computed corresponding bivariate correlations for each participant. As a result, for all but 1

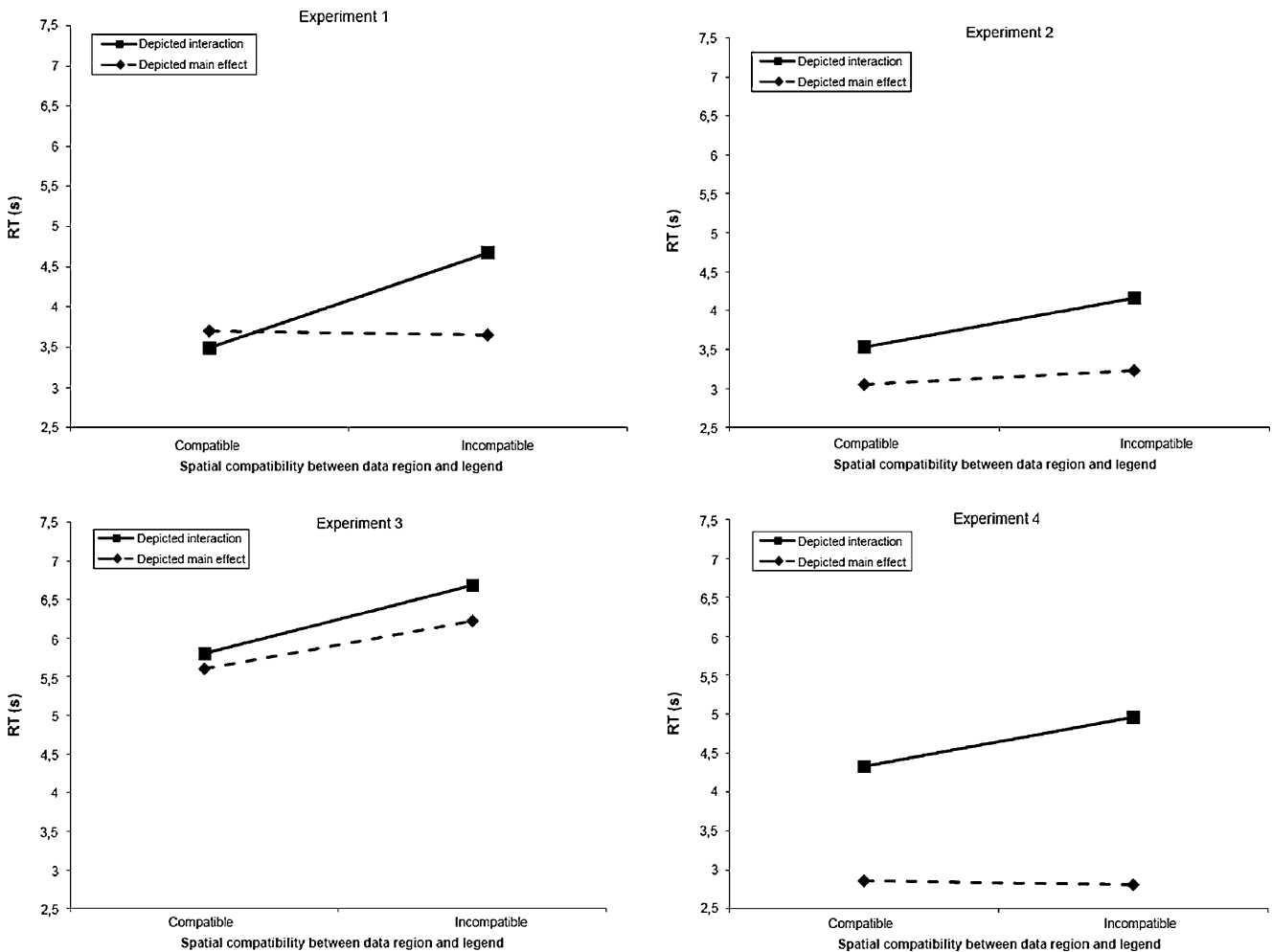


Fig. 2 Response times (RTs, in seconds) across experiments as a function of the spatial compatibility between the data region and the legend (compatible vs. incompatible) for depicted interactions and

main effects. Note that RTs in **Experiment 4** refer to statement validation times with respect to previously presented graphs

participant, we found significant positive correlations between the number of gaze transitions and RTs [mean $r = .55$, $CI_{.95} = .55 < r(63) < .67$].

Experiment 2 (line graphs) Outliers were defined as in Experiment 1 but did not occur throughout the data. RTs were greater for incompatible legends than for compatible legends, $F(1, 15) = 74.7$, $p < .001$, $\eta_p^2 = .83$ (see Fig. 2). There was no significant main effect of depicted effect type, $F(1, 15) = 2.9$, $p = .110$, but there was a significant interaction of depicted effect type and compatibility, $F(1, 15) = 43.2$, $p < .001$, $\eta_p^2 = .74$, which was of the same form as in Experiment 1. Posthoc t -tests revealed that there was no compatibility effect for depicted main effects (compatible, 3.70 s, $SE = 0.28$; incompatible, 3.65 s, $SE = 0.30$), $t < 1$, but there was for interactions (compatible, $M = 3.49$ s, $SE = 0.29$; incompatible, $M = 4.67$ s, $SE = 0.32$), $t(15) = 9.31$, $p < .001$.

Accuracy amounted to 95.65% ($SE = 1.00$) for compatible graphs, as compared with 93.55% ($SE = 1.50$) for incompatible graphs. However, this tendency yielded no significant effect of compatibility, $F(1, 15) = 2.76$, $p = .118$. There was neither a significant effect of depicted effect type, $F < 1$, nor a significant interaction, $F < 1$.

Gaze transitions between data region and legend were more frequent in incompatible ($M = 4.37$) than in compatible ($M = 3.63$) graphs, $F(1, 15) = 31.82$, $p < .001$, $\eta_p^2 = .68$. There was no main effect of the depicted effect type, $F(1, 15) = 1.72$, $p = .210$, but there was a significant interaction between depicted effect type and compatibility, $F(1, 15) = 6.63$, $p = .021$, $\eta_p^2 = .31$. Posthoc t -tests revealed that there was a significant difference between compatible and incompatible legends when participants had to judge a main effect ($M = 3.70$ gaze transitions, $SE = 0.28$, vs. $M = 4.04$ gaze transitions, $SE = 0.30$), $t(15) = 1.95$, $p = .035$, as well as for the interpretation of interactions ($M = 3.56$ gaze transitions, $SE = 0.32$, vs. $M = 4.69$ gaze transitions, $SE = 0.42$), $t(15) = 4.95$, $p < .001$, but the difference was larger for the depicted interactions. Overall, gaze transitions from legend to data region occurred more often ($M = 2.08$, $SE = 0.14$) than gaze transitions from data region to legend ($M = 1.92$, $SE = 0.16$), $t(15) = 2.56$, $p = .022$. The mean percentage of trials on which the legend was fixated before the data region across participants amounted to 83.6% ($SE = 2.04$). For all participants, we found significant positive correlations between the number of gaze transitions and RTs (mean $r = .64$, $CI_{.95} = .51 < r(63) < .74$).

A between-experiments comparison revealed that accuracy was greater for line graphs than for bar graphs for depicted main effects, $t(32) = 2.5$, $p = .02$, and for depicted interactions, $t(32) = 4.3$, $p < .001$, whereas RTs did not differ significantly (both t s < 1). Overall, gaze transitions were more frequent in Experiment 1 (bar graphs), $t(32) = 5.3$, $p < .001$.

Discussion

The results of Experiments 1 and 2 clearly revealed that compatible graphs were judged more quickly, more accurately (Experiment 1), and with fewer gaze transitions between data region and legend, as compared with incompatible graphs. When gaze transitions represent a marker for integration processes (Carpenter & Shah, 1998), this indicates that legend incompatibility indeed increased the difficulty of integrating relevant information from the data and the legend region, finally slowing down RTs. This claim is further corroborated by the observation that RTs were positively correlated with the amount of gaze transitions.

A closer inspection of the RT data revealed that the compatibility effect was increased (and sometimes only present) for the judgment of interactions, suggesting that a certain threshold of data pattern complexity and a respective increase of cognitive processing demands (Halford et al., 1998) needs to be crossed to achieve an effect of compatibility on performance in terms of RTs and/or accuracy. To further examine the influence of complexity, we conducted Experiment 3, in which more complex line graphs (consisting of more lines and more legend entries) were used.

Previous guidelines have recommended bar graphs for depicting differences between means across different levels of the independent variable and when the latter represents an ordinal or categorical variable (Gillan et al., 1998). However, our between-experiment comparison suggests a disadvantage for bar graphs even for the extraction of main effects in basic 2×2 designs. Probably, bar graphs generally produce too much visual clutter, and their use should be more limited than was previously assumed. The overall smaller number of gaze transitions in Experiment 1 may indicate that decoding of bar graphs was probably so demanding that participants tended to avoid rechecking the legend before responding, leading to poor overall performance. Alternatively, participants might have been less sensitive to the trial-by-trial change in the legends because of a lower salience of legends in bar graphs, due to the overall higher amount of “ink.”

Experiment 3

In Experiment 3, we explored to what extent the compatibility effect is modulated in graphs with greater visual complexity, where it is difficult to avoid compatibility issues by labeling lines directly without causing visual cluttering of textual elements. On the one hand, the relative spatial position of legend entries in complex graphs might be less salient, as compared with graphs with only two entries, so that

incompatibility might no longer play a major role. On the other hand, graphs with greater visual complexity should draw on more cognitive resources. If this enhances the potential for compatibility effects, one might expect RT effects not only for depicted interactions, but also for main effects. Since bar graphs seemed rather unsuitable for depicting interactions (see above), we focused only on line graphs.

Method

Participants Eighteen new right-handed students at RWTH Aachen University (16 of them female and 2 male) with a mean age of 21 years (range: 19–27) and normal or corrected-to-normal vision participated in the experiment.

Stimuli, apparatus, task, procedure, and design The graphs depicted the popularity of six school subjects for 6th and 9th graders (see Fig. 1). Popularity served as a dependent variable (y -axis), grade (6th vs. 9th) was depicted on the x -axis, and the six different school subjects were represented as black (uncrossed) lines defined in the legend on the right side.

Data point markers were of varying shape (triangle, square, or circle) and color (black or white). The legend contained six lines, each consisting of a marker and a label. Half of the lines were steeper than the remainder, so that corresponding statements could refer to both depicted main effects and interactions. Consequently, each graph was coupled with a statement referring to a main effect on one trial (e.g., “The overall popularity is greater for the subject physics as compared with religion”) and with a statement referring to an interaction on another trial (e.g., “The popularity increases more strongly from 6th to 9th grade for the subject mathematics as compared with sports”).

We designed 8 different graphs with randomized positions for the steeper lines, the specific markers, and the related school subjects. Each of the graphs was coupled with both a spatially compatible and an incompatible legend ($4^\circ \times 6^\circ$). Each of the resulting 16 graphs was coupled with both a statement corresponding to a main effect and a statement corresponding to an interaction that either matched the graph or did not, resulting in 64 trials. On incompatible trials, the relative spatial position of the two legend entries referred to by the statement was always incompatible with the corresponding spatial arrangement in the data region, and the remaining legend entries were randomly assigned to positions (positions compatible with the data region were excluded). Apparatus, task, procedure, and design were exactly the same as in [Experiment 2](#).

Results

Outliers were defined as in the previous experiments but did not occur throughout the data. RTs were greater for

incompatible legends ($M = 6.45$ s) than for compatible legends ($M = 5.70$ s), $F(1, 17) = 58.12$, $p < .001$, $\eta_p^2 = .77$ (see Fig. 2). There was no main effect of effect type, $F(1, 17) = 2.98$, $p > .10$, and no interaction of effect type and compatibility, $F(1, 17) = 1.74$, $p > .10$. Posthoc t -tests revealed a significant compatibility effect for depicted main effects, $t(17) = 4.11$, $p = .001$, and for interactions, $t(17) = 7.02$, $p < .001$.

Accuracy amounted to 95.5% ($SE = 0.90$) for compatible graphs and 94.5% ($SE = 1.10$) for incompatible graphs, $F < 1$. Accuracy was similar for the judgment of main effects ($M = 96.3\%$, $SE = 1.20$) and for interactions ($M = 96.0\%$, $SE = 0.90$), $F(1, 17) = 2.22$, $p > .10$. There was no significant interaction of effect type and compatibility, $F(1, 17) = 1.89$, $p > .10$.

Gaze transitions were significantly more frequent in incompatible ($M = 4.41$) than in compatible ($M = 4.13$) graphs, $F(1, 17) = 7.30$, $p = .015$, $\eta_p^2 = .30$. There was no main effect of the effect type, $F(1, 17) = 1.29$, $p > .10$, and no significant interaction between effect type and compatibility, $F < 1$. Overall, gaze transitions from legend to data region occurred more often ($M = 2.36$, $SE = 0.12$) than gaze transitions from data region to legend ($M = 1.90$, $SE = 0.09$), $t(17) = 7.21$, $p < .001$. The mean percentage of trials on which the legend was fixated before the data region across participants amounted to 94.3% ($SE = 1.47$). For all participants, we found significant correlations between the amount of gaze transitions and RTs (mean $r = .64$, $CI_{.95} = .55 < r(63) < .73$).

Discussion

[Experiment 3](#) demonstrated that the compatibility effect on RTs and gaze transitions generalizes to more visually complex graphs. Interestingly, we now found a significant compatibility effect on RTs, not only for depicted interactions, but also for main effects, suggesting that visual complexity plays a major role. Probably, the additional task of selecting the relevant pieces of data from the overwhelming amount of information in the data and legend regions increased overall processing demands, subsequently increasing adverse effects of incompatibility. The overall number of gaze transitions was in a similar range as in [Experiment 2](#), suggesting that in complex graphs, the number of relevant legend entries is more influential for legend revisitations than is their overall number.

Experiment 4

In [Experiment 4](#), we asked whether the compatibility effect transfers to a situation where relevant graph information is no longer visually present and, thus, needs to be retrieved

from memory. This corresponds to situations where we are exposed to graphs only for a limited amount of time—for example, while watching television or during a conference presentation. Thus, **Experiment 4** increased the complexity of task requirements by having participants judge graphs from memory. We used the same graphs and statements as in **Experiment 2** but reversed their within-trial presentation order. More specifically, participants were asked to judge whether a statement matched a graph that was previously presented for a limited amount of time.

Method

Participants Twelve new right-handed students at RWTH Aachen University (10 of them female and 2 male) with a mean age of 22 years (range: 19–27) and normal or corrected-to-normal vision participated in the experiment.

Stimuli, apparatus, task, procedure, and design The stimuli, apparatus, and design were the same as in **Experiment 2**. However, on each trial, the line graph was presented first for a fixed time interval of 5 s, preceded by the fixation cross (250 ms). The 5-s interval was based on the results from **Experiment 2**, where mean RTs did not exceed 5 s in any condition. Afterward, the statement was displayed, and participants judged as quickly and accurately as possible whether it corresponded to the previous graph. Then the fixation cross of the next trial appeared. Note that RTs here corresponded to statement verification times.

Results

Outliers were defined as in the previous experiments but did not occur throughout the data. RTs for judging statements corresponding to graphs with incompatible legends were greater than those for statements corresponding to graphs with compatible legends, $F(1, 11) = 8.56$, $p = .014$, $\eta_p^2 = .44$ (see Fig. 2). The main effect of depicted effect type was also significant, $F(1, 11) = 16.48$, $p = .002$, $\eta_p^2 = .60$. Furthermore, there was a significant interaction of depicted effect type and compatibility, $F(1, 11) = 9.76$, $p = .010$, $\eta_p^2 = .47$. Posthoc *t*-tests revealed no significant difference between compatible and incompatible legends for the judgment of main effects ($M = 2.86$ s, $SE = 0.28$ vs. $M = 2.81$ s, $SE = 0.26$), $t < 1$, but a significant difference for the judgment of interactions ($M = 4.33$ s, $SE = 0.53$ vs. $M = 4.96$ s, $SE = 0.60$), $t(11) = 4.12$, $p = .002$.

Accuracy amounted to 90.1% ($SE = 1.60$) for incompatible graphs and 91.7% ($SE = 2.10$) for compatible graphs, $F(1, 11) = 1.07$, $p = .327$. Accuracy was greater for the verification of main effects ($M = 93.2\%$, $SE = 1.80$), as compared with interactions ($M = 88.5\%$, $SE = 2.10$), $F(1, 11) = 6.91$, $p = .023$, $\eta_p^2 = .39$. Additionally, there was

a significant interaction of depicted effect type and compatibility, $F(1, 11) = 5.33$, $p = .041$, $\eta_p^2 = .33$. Post hoc *t*-tests revealed no significant compatibility effect for the judgment of main effects ($M = 94.3\%$, $SE = 1.40$, for incompatible graphs vs. $M = 92.2\%$, $SE = 2.40$, for compatible graphs), $t < 1$, but there was for interactions ($M = 85.9\%$, $SE = 2.60$, for incompatible graphs vs. $M = 91.1\%$, $SE = 2.20$, for compatible graphs), $t(11) = 2.16$, $p = .027$.

Whereas RTs and accuracy data were related to the judgment of the statements, the analysis of eye movements here refers to the visual processing of the graphs that were presented prior to each statement. As in all the previous experiments, gaze transitions between data region and legend were significantly more frequent in incompatible ($M = 4.49$) than in compatible ($M = 4.18$) graphs, $F(1, 11) = 7.46$, $p < .020$, $\eta_p^2 = .40$. There was no main effect of the depicted effect type, $F < 1$, and no significant interaction between depicted effect type and compatibility, $F(1, 11) = 1.48$, $p = .249$. A further analysis revealed that, overall, gaze transitions from legend to data region tended to occur more often ($M = 2.22$, $SE = 0.18$) than gaze transitions from data region to legend ($M = 2.10$, $SE = 0.17$), although this difference did not reach statistical significance, $t(11) = 1.75$, $p = .10$. The mean percentage of trials on which the legend was fixated before the data region across participants amounted to 87.0% ($SE = 2.64$). For 9 participants, we did not find significant correlations between the number of gaze transitions and RTs in the subsequent statement validation task; the remainder consisted of one significant negative and two significant positive correlations (mean $r = .07$, $CI_{.95} = -.04 < r(63) < .19$). Taken together, gaze transitions and RTs were not consistently related.

Discussion

Experiment 4 revealed adverse effects of legend incompatibility (significant RT effect for depicted interactions, but not for main effects) even under complex task requirements—that is, when graphs are judged from memory. One might have expected no compatibility effect at all on statement verification times in the present setting: Data–legend incompatibility, which does not affect the meaning of a graph, could be considered as surface information, which, in turn, is known to decay rapidly (Kintsch et al., 1990) and, thus, should no longer be available when the statements are read. However, this assumption can be rejected, since the data clearly showed that legend incompatibility negatively affected subsequent statement verification times. Probably, participants stored an impoverished abstract visual representation of the graph in working memory, which was further processed once the statement defined relevant elements that needed to be integrated. During statement processing, they might

have selectively attended to the relevant elements of the visual representation in working memory, leaving room for a compatibility effect to occur. Performance would then have been determined by attention processes during visual retention, which may be regarded as covert mental attention shifts related to a visual memory representation, analogous to the overt gaze transitions observed in the previous experiments.

The finding that statements referring to interactions were processed more slowly than those referring to main effects most probably reflects the fact that interaction statements were longer and, thus, take longer to read. However, it is interesting to note that during the judgment of interactions, the post hoc tests additionally revealed evidence for reduced accuracy for the judgment of incompatible graphs, as compared with compatible graphs. Probably, the longer reading times of the interaction statements went hand in hand with a deterioration of the visual memory representation of the graph, thus increasing the difficulty of executing covert mental attention shifts to update relevant information from the legend region in incompatible graphs.

General discussion

The aim of the present study was to analyze integration processes in graph comprehension by systematically manipulating spatial compatibility of elements in the legend and the data region in tasks of varying complexity. Participants in Experiments 1, 2, and 3 judged whether bar or line graphs corresponded to previously presented statements. Most important, the results, overall, supported the hypothesis that incompatible data–legend relations can slow down RTs (Experiments 1, 2 and 3) and increase errors (Experiment 1), even though the spatial position of legend entries was irrelevant for the purpose of the task. Experiment 4 extended the theoretical significance of this compatibility effect by demonstrating that incompatible graphs also affected information retrieval from memory in the absence of direct visual stimulation.

The influence of spatial compatibility on graph comprehension corresponds to at least two well-known display design principles—namely, the principle of congruence and the proximity compatibility principle (Wickens & Carswell, 1995; Wickens & Hollands, 2000). The principle of congruence was originally applied to congruence between instructional order and temporal procedures ordering and to stimulus ordering and response ordering. However, the present data suggest that it appears perfectly applicable to graph element and legend ordering. The proximity compatibility principle implies that similarity (perceptual proximity) of graph elements promotes their integration (processing proximity). The concept of similarity, which

was already introduced as a Gestalt principle in the seminal work of Max Wertheimer (1923; see also Ware, 2004), can refer to perceptual attributes such as color and texture, but also to spatial position (see Wickens & Hollands, 2000). More specifically, Gillan et al. (1998) explicitly predicted that the ordering of legend symbols should match the order of indicators in the data region, a claim that is directly supported by our data.

Our present results bear interesting implications for the proximity compatibility principle, since they demonstrate that display proximity can also take the form of *relative* spatial proximity (referring to the relative position of elements in the legend and data region), thus extending the original claim that (absolute) spatial proximity plays a crucial role. However, it appears noteworthy that an advantage of the compatible conditions was not observed throughout all the experimental conditions. Thus, the proximity compatibility principle does not, per se, hold in any given situation, but only in tasks of sufficient (data pattern) complexity that a certain amount of cognitive load is induced. Overall, this conclusion is well in line with previous predictions made by Wickens and Carswell (1995), who substantiated their assumption that close physical proximity serves integration processes by referring to potential problems related to interference between information access (resulting in movements of attention to the different sources of information) and cognitive complexity of the mental integration operation (resulting in higher working memory load). These assumptions seem to be confirmed by the present set of data: Information access costs were directly created by a lack of data–legend compatibility and were reflected in the gaze patterns. These problems regarding information access, in turn, interfered with visual complexity and the complexity of the data pattern, where higher levels created (or amplified) congruence costs.

Since in Experiments 1 and 2, the compatibility effect on RTs occurred only for the judgment of interactions, and not for main effects, *data pattern complexity* seemed to play a major role for its occurrence (see Halford et al., 1998). In Experiment 3, the compatibility effect was also present for the judgment of depicted main effects, showing that *visual complexity* effectively modulated the compatibility effect, too. However, the increase in the complexity of task requirements in Experiment 4 yielded a similar data pattern as in the otherwise comparable Experiment 2, suggesting that the specific task requirements had no sizable impact. If we interpret the compatibility effect as an empirical marker for integration processes (see below), we can confirm that integrative processes tend to scale up as complexity increases (Ratwani et al., 2008). However, the present data further specify this claim by showing that this tendency seems to be limited to visual and data pattern complexity

and does not hold in the same way for complexity of task requirements (at least, as it was implemented in the present study).

Overall, the compatibility effect further specifies previous theoretical assumptions proposed by Carpenter and Shah (1998). They emphasized the importance of integration processes during graph comprehension, during which the meaning of data is inferred by combining information from the data region and the legend (i.e., “tracking the referents”). More specifically, they stated that this process should be reflected in frequent gaze transitions between these graph elements. However, in their framework, it remained an open question as to exactly which factors and mechanisms mediate the efficiency of integration processes. On the basis of the present study, we demonstrated that spatial compatibility in combination with task complexity is a major determinant of this process.

Specifying the mechanisms of data–legend integration All four experiments consistently showed that gaze transitions between the legend and the data region were more frequent for incompatible data–legend relations. Additional intra-individual correlation analyses confirmed that the number of gaze transitions was correlated with RTs, suggesting that the integration of elements and their meanings is, indeed, a crucial factor in determining the efficiency of understanding a graph. Note, however, that the corresponding correlations were far from perfect, which, in some instances, led to differential result patterns for RTs and gaze transitions. For example, in [Experiment 1](#), there was a significant interaction between compatibility and the depicted effect type on RTs, which did not show up in gaze transitions. This demonstrates that the eye movements (and the corresponding cognitive processes) are not the sole source of variance that finally determines RTs. For example, it is possible that a time loss associated with additional eye movements can, to some extent, be compensated through parallel cognitive processing, so that an increase of oculomotor effort does not necessarily result in substantially delayed decision times.

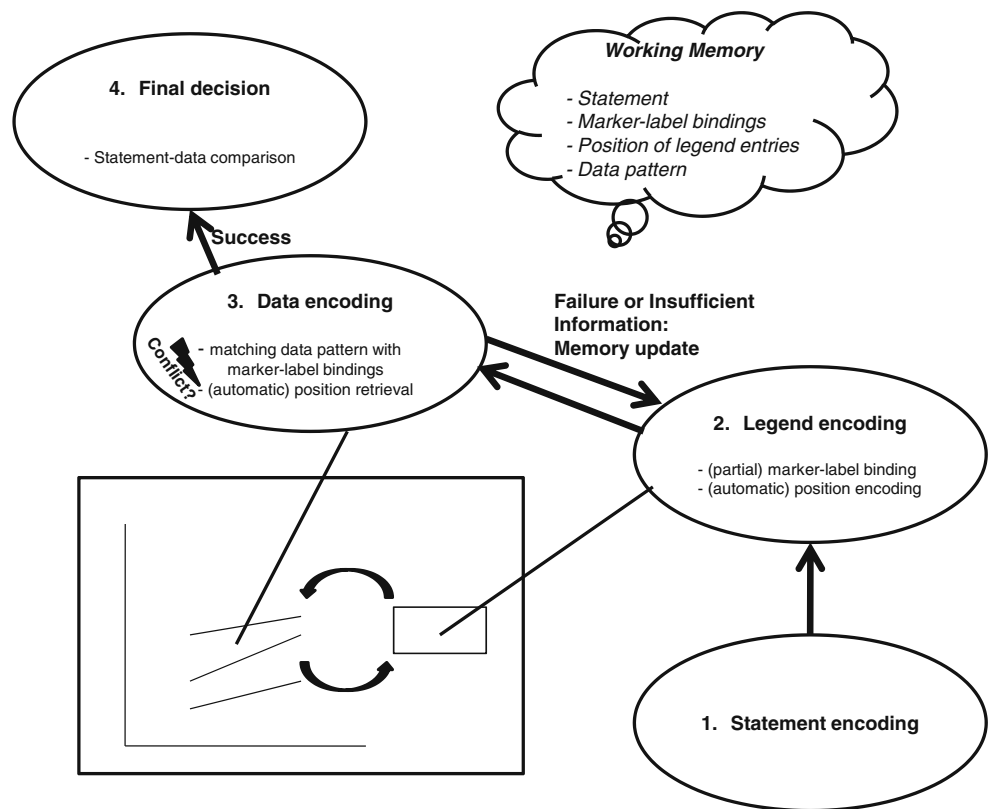
Some more recent models of graph comprehension explicitly model corresponding scan paths that occur during specific procedural steps required to perform a graph comprehension task (e.g., the graph-based reasoning model; Peebles & Cheng, 2002). Interestingly, this line of research has revealed that the assumption of an optimal scan path (including necessary transitions only) may gloss over important strategic and other cognitive factors at the participant’s level (Peebles & Cheng, 2003). More specifically, Peebles & Cheng (2003) reported data showing that the amount of gaze transitions between parts of a question and a corresponding graph exceeded the number of necessary transitions, based on formal task analyses. This indicates that although a certain amount of gaze transitions

reflect necessary procedural steps during the integration phase in a graph comprehension task, other (additional) transitions may result from workingmemory limitations (forgetting) or may be based on strategic factors at the participant’s level—for example, the choice to delay encoding selected legend information until it is required, in turn triggering rereading of individual legend elements during the processing of the graph (Peebles & Cheng, 2003).

Figure 3 depicts a visualization of some of the potential mechanisms in data–legend integration within our present study, starting with statement encoding. Eye-scanning routines often started with a legend fixation, triggering the encoding of the relation between markers and their meaning, before the data region was entered. The compatibility effect suggests that during the inspection of the legends, participants did not relate the verbal label solely to the marker information, but also, at least on some trials, to its (task-irrelevant) spatial position. Probably, this task-irrelevant information automatically generates the expectation that the spatial layout in the data region is designed accordingly. This expectation might yield a strategic benefit for subsequent processing whenever the expectation is met, but its violation through incompatible elements in the data region may push working memory capacities to their limits (especially in complex tasks), triggering legend revisitations to update memory. Thus, the increase of revisitations to the legend in incompatible graphs may suggest that the adverse effects of incompatibility were due mainly to interference or capacity limits in working memory, and not only to a more time-consuming process of memory retrieval of legend information during the fixation of the data region. Additionally, some revisitations to the legend may also be due to strategic factors—for example, the choice to delay encoding selected legend information until it is required, which may trigger rereading of certain legend elements (Peebles & Cheng, 2003). Overall, these mechanisms further specify the notion of “tracking the referents” during integration processes (Carpenter & Shah, 1998; Peebles & Cheng, 2003; Ratwani et al., 2008).

Limitations and implications Although many empirically informed guidelines for display design exist (for reviews, see Cleveland, 1985; Hitt, Schutz, Christner, Ray, & Coffey, 1961; Kosslyn, 1989, 1994; Meyer, 2000; Pinker, 1990; Shah & Hoefner, 2002), some issues remained controversial (Coll & Coll, 1993; De Sanctis, 1984; Jarvenpaa, Dickson, & De Sanctis, 1985) or have not been addressed yet, including the role of data–legend compatibility. Here, we present the first empirical evidence for the advice to use compatible legends, especially when readers are supposed to extract interactions. [Experiment 4](#) showed that even memory-based decisions suffer from legend incompatibility. Taken together, the adverse effect of data–legend incompat-

Fig. 3 Visualization of potential data–legend integration processes discussed in the General Discussion section (see the General Discussion for further details)



ibility appears to be a robust and reliable phenomenon, and there is no reason to doubt that it also transfers to other (e.g., more applied) settings in a similar way.

Despite the general advice to use compatible legends, some limitations should be addressed. First, in the present experiments, there was always an unambiguous solution for a compatible legend arrangement, whereas in some graphs, this may not be possible (e.g., disordinal interactions involving line crossings). Although the present data do not allow us to directly infer design guidelines for graphs including crossed lines, the results of previous studies, along with the present data, strongly suggest that legend entries in line graphs (with legends on the right side) should be ordered in the same way as the rightmost endpoints of the lines. First, Parkin (1983; in Pinker, 1990) showed that the Gestalt principle of proximity plays a major role during graph comprehension, suggesting that legend entries should be proximal to the respective lines. If legends are used in the first place, this Gestalt principle is best instantiated when the sequence of legend entries corresponds to the proximal pattern of lines—that is, their rightmost endpoints (for legends on the right). Second, the present data showed more frequent gaze transitions from the legend to the data region than vice versa. This strongly indicates that participants start by encoding the legend, rather than reading a graph from left to right.

A further limitation of the present study refers to the fact that we examined only single graphs within one experi-

mental trial. However, when several graphs are based on the same set of variables but depict different data, consistent-spatial arrangements (probably following a semantically meaningful sequence—e.g., medication doses) across graphs may outweigh adverse effects of within-graph incompatibility (Andre & Wickens, 1992). These research questions clearly require closer attention in future research.

Conclusions In sum, the present study showed that legend compatibility affects integration processes during graph comprehension but that this effect is modulated by concurrent processing demands defined through different types of complexity. Adverse effects of incompatibility were accompanied by an increase of gaze transitions between the data region and the legend, suggesting that legend revisitations served to counteract forgetting in working memory. Surprisingly, the compatibility effect could be replicated even when relevant graph information needed to be retrieved from memory, indicating that integration plays a major role even in the absence of direct visual stimulation.

Author Note Lynn Huestegge and Andrea M. Philipp, Institute for Psychology, RWTH Aachen University, Aachen, Germany. We thank Iring Koch for the study's initial idea and those who kindly volunteered to participate. Additionally, we thank Yaffa Yeshurun, Chris Wickens, and an anonymous reviewer for helpful comments on an earlier draft of the manuscript.

References

- American Psychological Association. (2006). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.
- American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.
- Andre, A. D., & Wickens, C. D. (1992). Compatibility and consistency in display-control systems: Implications for aircraft decision and design. *Human Factors*, *34*, 639–653.
- Bryant, D. J., & Tversky, B. (1999). Mental representations of perspective and spatial relations from diagrams and models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 137–156.
- 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.
- Carswell, C. M., Frankenberger, S., & Bernhard, D. (1991). Graphing in depth: Perspectives on the use of three-dimensional graphs to represent lower dimensional data. *Behaviour and Information Technology*, *10*, 459–474.
- Carswell, C. M., & Wickens, C. D. (1990). The perceptual integration of graphical attributes: Configurality, stimulus homogeneity, and object integration. *Perception & Psychophysics*, *47*, 157–168.
- Cleveland, W. S. (1985). *The elements of graphing data*. Monterey: Wadsworth.
- Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, *77*, 541–547.
- Cleveland, W. S., & McGill, R. (1986). An experiment in graphical perception. *International Journal of Man–Machine Studies*, *25*, 491–500.
- Cleveland, W. S., & McGill, R. (1987). Graphical perception: The visual decoding of quantitative information on graphical displays of data. *Journal of the Royal Statistical Society A*, *150*, 192–229.
- Coll, J. H., & Coll, R. (1993). Tables and graphs: A classification scheme for display presentation variables and a framework for research in this area. *Information Processing and Management*, *29*, 745–750.
- De Sanctis, G. (1984). Computer graphics as decision aids: Directions for research. *Decision Sciences*, *15*, 463–487.
- Feeny, A., Hola, A. K. W., Liversedge, S. P., Findlay, J. M., & Metcalfe, R. (2000). How people extract information from graphs: Evidence from a sentence-graph verification paradigm. *Theory and Application of Diagrams, Proceedings*, 149–161.
- Fischer, M. H. (2000). Do irrelevant depth cues affect the comprehension of bar graphs? *Applied Cognitive Psychology*, *14*, 151–162.
- Fischer, M. H., Dewulf, N., & Hill, R. L. (2005). Designing bar graphs: Orientation matters. *Applied Cognitive Psychology*, *19*, 953–962.
- Fitts, P. M., & Simon, C. W. (1952). Some relations between stimulus patterns and performance in a continuous dual-pursuit task. *Journal of Experimental Psychology*, *43*, 428–436.
- 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.
- Gillan, D. J. (1995). Visual arithmetic, computational graphics, and the spatial metaphor. *Human Factors*, *37*, 766–780.
- Gillan, D. J., & Lewis, R. (1994). A componential model of human interaction with graphs: I. Linear regression modelling. *Human Factors*, *36*, 419–440.
- Gillan, D. J., Wickens, C. D., Hollands, J. G., & Carswell, C. M. (1998). Guidelines for presenting quantitative data in HFES publications. *Human Factors*, *40*, 28–41.
- Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *The Behavioral and Brain Sciences*, *21*, 803–865.
- Harris, R. L. (1999). *Information graphics: A comprehensive illustrated reference*. New York: Oxford University Press.
- Hitt, W. D., Schutz, H. G., Christner, C. A., Ray, H. W., & Coffey, L. J. (1961). Development of design criteria for intelligence display formats. *Human Factors*, *3*, 86–92.
- Hommel, B., & Prinz, W. (1997). *Theoretical issues in stimulus–response compatibility* (Vol. 119). Amsterdam: Elsevier.
- Jarvenpaa, S. L., Dickson, G. W., & De Sanctis, G. (1985). Methodological issues in experimental IS research: Experiences and recommendations. *MIS Quarterly*, *9*, 141–156.
- Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language*, *29*, 133–159.
- Kornblum, S. (1992). Dimensional overlap and dimensional relevance in stimulus–response and stimulus–stimulus compatibility. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior* (Vol. II, pp. 743–777). Amsterdam: North-Holland.
- Kosslyn, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology*, *3*, 185–225.
- Kosslyn, S. M. (1994). *Elements of graph design*. New York: Freeman.
- Larkin, J. A., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, *11*, 65–99.
- Lohse, J. (1991). *A cognitive model for the perception and understanding of graphs*. In Proceedings of the SIGHI Conference on Human Factors in Computing Systems: Reaching through technology (pp. 137–144). New York: Association for Computing Machinery.
- Lohse, G. L. (1993). A cognitive model for understanding graphical perception. *Human–Computer Interaction*, *8*, 353–388.
- Meyer, J. (2000). Performance with tables and graphs: effects of training and a Visual Search Model. *Ergonomics*, *11*, 184–186.
- Meyer, J., Shinar, D., & Leiser, D. (1997). Multiple factors that determine performance with tables and graphs. *Human Factors*, *39*, 268–286.
- Peebles, D. (2008). The effect of emergent features on judgments of quantity in configural and separable displays. *Journal of Experimental Psychology: Applied*, *14*, 85–100.
- Peebles, D., & Cheng, P. C.-H. (2001). Extending task analytic models of graph-based reasoning: A cognitive model of problem solving with Cartesian graphs in ACT-R/PM. In *Proceedings of the Fourth International Conference on Cognitive Modeling* (pp. 169–174). Mahwah, NJ: Erlbaum.
- Peebles, D., & Cheng, P. C.-H. (2002). Extending task analytic models of graph-based reasoning: A cognitive model of problem solving with Cartesian graphs in ACT-R/PM. *Cognitive Systems Research*, *3*, 77–86.
- 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*, 28–46.
- Pinker, S. (1990). A theory of graph comprehension. In R. Freedle (Ed.), *Artificial intelligence and the future of testing* (pp. 73–126). Hillsdale: Erlbaum.
- Proctor, R. W., & Vu, K.-P. L. (2006). *Stimulus–response compatibility principles: Data, theory, and application*. Boca Raton: Taylor & Francis.
- Ratwani, R. M., Trafton, J. G., & Boehm-Davis, D. A. (2008). Thinking graphically: Connecting vision and cognition during

- graph comprehension. *Journal of Experimental Psychology: Applied*, 14, 36–49.
- Renshaw, J. A., Finlay, J. E., Tyfa, D., & Ward, R. D. (2004). Understanding visual influence in graph design through temporal and spatial eye movement characteristics. *Interacting with Computers*, 16, 557–578.
- Schmid, C. F. (1983). *Statistical graphics: Design principles and practices*. New York: Wiley.
- Shah, P. (1995). *Cognitive processes in graph comprehension. Unpublished doctoral dissertation*. Pittsburgh: Carnegie Mellon University.
- Shah, P., & Carpenter, P. A. (1995). Conceptual limitations in comprehending line graphs. *Journal of Experimental Psychology: General*, 124, 43–61.
- Shah, P., & Hoeffner, J. (2002). Review of graph comprehension research: Implications for instruction. *Educational Psychology Review*, 14, 47–69.
- Siegrist, M. (1996). The use or misuse of three-dimensional graphs to represent lower-dimensional data. *Behaviour and Information Technology*, 15, 96–100.
- Simkin, D., & Hastie, R. (1987). An information-processing analysis of graph perception. *Journal of the American Statistical Association*, 82, 454–465.
- Sparrow, J. A. (1989). Graphical displays in information systems: Some data properties influencing the effectiveness of alternative formats. *Behaviour and Information Technology*, 8, 43–56.
- Spence, I. (1990). Visual psychophysics of simple graphical elements. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 683–692.
- Tufte, E. R. (1983). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.
- Vessey, I. (1991). Cognitive fit: A theory-based analysis of the graphs versus tables literature. *Decision Sciences*, 22, 219–240.
- Vessey, I. (1994). The effect of information presentation on decision making: A cost–benefit analysis. *Information Management*, 27, 103–119.
- Ware, C. (2004). *Information visualisation: Perception for design* (2nd ed.). San Francisco: Morgan Kaufman.
- Washburne, J. N. (1927). An experimental study of various graphic, tabular and textual methods of presenting quantitative material. *Journal of Educational Psychology*, 18, 465–476.
- Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt. *Psychologische Forschung: Zeitschrift für Psychologie und ihre Grenzwissenschaften*, 4, 301–350.
- Wickens, C. D., & Andre, A. D. (1990). Proximity compatibility and information display: Effects of color, space and objectness on information integration. *Human Factors*, 32, 61–78.
- Wickens, C. D., & Carswell, C. M. (1995). The proximity compatibility principle: Its psychological foundation and its relevance to display design. *Human Factors*, 37, 473–494.
- Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance* (3rd ed.). Upper Saddle River, NJ: PrenticeHall.
- Zacks, J., Levy, E., Tversky, B., & Schiano, D. J. (1998). Reading bar graphs: Effects of depth cues and graphical context. *Journal of Experimental Psychology: Applied*, 4, 119–138.
- Zacks, J., & Tversky, B. (1999). Bars and lines: A study of graphic communication. *Memory & Cognition*, 27, 1073–1097.