

**Thinking with data visualisations: cognitive processing and spatial inferences
when communicating climate change**

Jordan Philip Harold

100068596

Submitted for consideration for a Doctorate of Philosophy in Psychology

University of East Anglia, School of Psychology

September, 2017

© This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived there from must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

Abstract

Data visualisations can be effective for communicating scientific data, but only if they are understood. Such visualisations (i.e. scientific figures) are used within assessment reports produced by the Intergovernmental Panel on Climate Change (IPCC). However, IPCC figures have been criticised for being inaccessible to non-experts. This thesis presents a thematic analysis of interviews with IPCC authors, finding that a requirement to uphold scientific accuracy results in complex figures that are difficult for non-experts to comprehend, and which therefore require expert explanation. Evidence is subsequently presented showing that figures with greater visual complexity are associated with greater perceived comprehension difficulty among non-experts. Comprehension of complex data visualisations may require readers to make spatial inferences. When interpreting a time-series graph of climate data, it was found that non-experts did not always readily identify the long-term trend. Two experiments then show that linguistic information in the form of warnings can support spatial representations for trends in memory by directing visual attention during encoding (measured using eye-tracking). This thesis also considers spatial inferences when forming expectations about future data, finding that expectations were sensitive to patterns in past data. Further, features that act on bottom-up perceptual processes were largely ineffective in *supporting* spatial inferences. Conversely, *replacing* spatial inferences by explicitly representing information moderated future expectations. However, replacing spatial inferences might not always be desirable in real-world contexts. The evidence indicates that when information is not explicitly represented in a data visualisation, providing top-down knowledge may be more effective in supporting spatial inferences than providing visual cues acting on bottom-up perceptual processes. This thesis further provides evidence-based guidelines drawn from the cognitive and psychological sciences to support climate change researchers in enhancing the ease of comprehension of their data visualisations, and so enable future IPCC outputs to be more accessible.

Contents

Abstract	2
List of Tables	6
List of Figures	10
Acknowledgements and declarations	14
Chapter 1: Introduction	15
Cognition for scientific data visualisations.....	18
Intuitions for effective data visualisations	21
Improving accessibility of scientific data visualisations	22
The role of visual attention	22
Directing attention by visual design.....	24
Directing attention by informing expectation	26
Handling complexity.....	26
Supporting inference-making.....	28
Using text to support cognition.....	29
Tailoring data visualisations to different audiences.....	31
Gaps in current knowledge	32
The purpose and outline of this thesis	34
Chapter 2: Are there comprehension difficulties with IPCC figures?.....	36
The IPCC - background context	37
Communication challenges with IPCC reports.....	37
IPCC data visualisations	39
Study 1: What factors influence the production and communication of IPCC figures?	41
Method	41
Analysis and results	46
Discussion	69
Study 2: Perceived ease of comprehension of IPCC figures	73
Method	75
Results.....	77
Discussion	82
Study 3: Visual complexity of IPCC Figures	85
What is visual complexity and how can it be measured?	85

Computational measures of visual clutter	87
Subband entropy as a measure of visual clutter	88
The present study / design.....	89
Method	89
Results	90
Discussion	92
General discussion	95
Chapter 3: Supporting spatial inferences.....	98
Interpreting trends in data	99
Perception of a complex line	101
Study 4 (pilot study): Do non-experts describe trends in an IPCC figure?	103
Method	103
Results	106
Discussion	107
Study 5: Can language support spatial inferences for trends?	109
Method	110
Results	117
Discussion	122
Study 6: What language supports spatial inferences for trends?	124
Method	125
Results	128
Discussion	136
General discussion	137
Chapter 4: Forming expectations from data	140
How do people form expectations about the future from time-series graphs?	141
Study 7: Expectations with and without a trend line	145
Method	146
Results	151
Discussion	160
Study 8: Expectations with and without an informational arrow	162
Method	163
Results	164
Discussion	171
Study 9: Expectations in horizontal and vertical planes	173

Method	174
Results	176
Discussion	182
General discussion	183
Chapter 5: Discussion.....	188
Results overview.....	189
Translating insights from cognitive and psychological research into practice	195
Putting guidance in to practice	200
Limitations	201
Future directions	204
Conclusions.....	205
References	207
Appendix 1	233
Interview protocol / guide.....	233
Appendix 2	238
Graph stimuli allocation to blocks for Study 5.....	238
Appendix 3	240
Data tables (means and standard deviations) for Study 5 analyses.	240
Appendix 4	242
Graph stimuli allocation to blocks for Study 6.....	242
Appendix 5	244
Data tables (means and standard deviations) for Study 6 analyses.	244
Appendix 6	248
Data tables (means and standard deviations) for Study 7 analyses.	248
Appendix 7	251
Data tables (means and standard deviations) for Study 8 analyses.	251
Appendix 8	254
Data tables (means and standard deviations) for Study 9 analyses.	254
Appendix 9	257
Harold, Lorenzoni, Shipley & Coventry (2016).....	257
Appendix 10	293
Harold, Coventry, Lorenzoni, & Shipley (2015).....	293

List of Tables

Table 1. Study 1 comparison of demographics of the interviewed sample to the full set of authors to the AR5 WG1 SPM.

Table 2. Study 1 definitions of themes and their sub-themes

Table 3. Study 4 coding criteria for the four characteristics.

Table 4. Study 4 frequency of the number of individuals who verbally described each characteristic.

Table 5. Study 5 mixed ANOVA table (test graph x x-ticks x warning); * indicates significance at the .05 level.

Table 6. Study 5 mixed ANOVA table (test graph x block x warning); * indicates significance at the .05 level.

Table 7: Study 5 mean (*M*) and standard deviations (*SD*) of fixation duration in ms during study for each AOI.

Table 8. Study 6 mixed ANOVA table (test graph x block x warning); * indicates significance at the .05 level.

Table 9. Meta analyses of the effect size of the paired difference between sensitivity for gradient differences and amplitude differences in Study 5 and 6; negative effect sizes indicate better performance on amplitude different trials, positive effect sizes indicate better performance on gradient different trials.

Table 10. Study 6 mixed ANOVA table (test graph x warning user goal x warning informational content); * indicates significance at the .05 level.

Table 11. Study 6 mixed ANOVA table (test graph x warning informational content); * indicates significance at the .05 level.

Table 12. Study 7 mixed ANOVA for changes in mean location of expected future values; * indicates significance at the .05 level. † indicates Greenhouse-Geisser corrected statistic.

Table 13. Study 7 mixed ANOVA for mean acceptance rates of the outer probes; * indicates significance at the .05 level.

Table 14. Study 7 mixed ANOVA for mean confidence ratings on ‘yes’ judgements; * indicates significance at the .05 level.

Table 15. Study 8 mixed ANOVA for changes in mean location of expected future values; * indicates significance at the .05 level. † indicates Greenhouse-Geisser corrected statistic.

Table 16. Study 8 mixed ANOVA for mean acceptance rates of the outer probes; * indicates significance at the .05 level.

Table 17. Study 8 mixed ANOVA for mean confidence ratings on ‘yes’ judgements; * indicates significance at the .05 level.

Table 18. Study 9 mixed ANOVA for changes in mean location of expected future values; * indicates significance at the .05 level.

Table 19. Study 9 mixed ANOVA for mean confidence ratings on ‘yes’ judgements; * indicates significance at the .05 level.

Table 20. Summary of key findings across studies 7-9.

Table 21. Evidence-informed guidelines to improve accessibility of scientific data visualisations of climate science.

Table A2-1. Study 5 same-different graph stimuli allocation to blocks.

Table A2-2. Study 5 filler trials graph stimuli allocation to blocks.

Table A3-1. Study 5 means and standard deviations of sensitivity (d') as a function of trial type, warning and x-ticks.

Table A3-2. Study 5 means and standard deviations of sensitivity (d') as a function of trial type, warning and block.

Table A4-1. Study 6 same-different graph stimuli allocation to blocks.

Table A4-2. Study 6 filler trials graph stimuli allocation to blocks.

Table A5-1. Study 6 means and standard deviations of sensitivity (d'), as a function of trial type and block, for the 'no warning' and 'identify trend, ignore extreme warning' groups.

Table A5-2. Study 6 means and standard deviations of sensitivity (d') on completely different trials as a function of block, warning user goal and warning informational content.

Table A5-3. Study 6 means and standard deviations of sensitivity (d') as a function of trial type, warning user goal and warning informational content.

Table A5-4. Study 6 means and standard deviations of sensitivity (d') as a function of trial type and warning informational content, including the 'no warning' group.

Table A6-1. Study 7 means and standard deviations of mean location of expected future values (in units of SDs of the distribution of the noise) as a function of recent data direction, trend line, global trend direction and block order.

Table A6-2. Study 7 means and standard deviations of the mean number of 'yes' responses to the outer probes as a function of trend line and block order.

Table A6-3. Study 7 means and standard deviations of VAS scores for trials in which the probe was judged consistent with past data, as a function of probe location, trend line and block order.

Table A7-1. Study 8 means and standard deviations of mean location of expected future values (in units of SDs of the distribution of the noise) as a function of recent data direction, arrow, global trend direction and block order.

Table A7-2. Study 8 means and standard deviations of the mean number of 'yes' responses to the outer probes as a function of arrow and block order.

Table A7-3. Study 8 means and standard deviations of VAS scores for trials in which the probe was judged consistent with past data, as a function of probe location, arrow and block order.

Table A8-1. Study 9 means and standard deviations of mean location of expected future values (in units of SDs of the distribution of the noise) as a function of recent data direction, orientation, global trend direction and block order.

Table A8-2. Study 9 means and standard deviations of the mean number of 'yes' responses to the outer probes as a function of orientation and block order.

Table A8-3. Study 9 means and standard deviations of VAS scores for trials in which the probe was judged consistent with past data, as a function of probe location, orientation and block order.

List of Figures

Figure 1. Figure SPM.5 from IPCC Fourth Assessment Report, Working Group 1, Summary for Policy Makers (IPCC, 2007).

Figure 2: Conceptual overview of the process of comprehension for data visualisations and approaches to improving accessibility.

Figure 3. Example of visual attention for an IPCC figure for a non-expert viewer trying to interpret the data visualisation (measured using eye tracking: first 15 seconds of data shown). a: eye gaze shown as individual fixations and connections between fixations; b: areas receiving visual attention; computed from the locations of the fixations, weighted by the duration of each fixation.

Figure 4. Schematic of properties known to direct visual attention that can be used in the design of data visualisations to help direct viewers' attention to important information.

Figure 5. Thumbnails images of the ten figures from the IPCC AR5 Working Group 1 Summary for Policy Makers (IPCC, 2013a). Larger versions of figures are not provided here due to copyright, but can be accessed in reference IPCC, 2013a.

Figure 6. Summary of themes and sub-themes.

Figure 7. Rank order of IPCC AR5 Working Group 1 SPM figures based on their perceived ease of comprehension across all undergraduates – figures are shown from easiest (rank 1) to most difficult (rank 10). Mean and standard deviation of the ranks are provided under each figure.

Figure 8. Rank order, across all experts, for IPCC AR5 Working Group 1 SPM figures based on their importance to inform future policy. Figures shown from most important (rank 1) to least important (rank 10). Mean and standard deviation of the ranks are provided under each figure.

Figure 9. Subband entropy scores for each figures, numbered from the least cluttered (lowest subband entropy score) through to the most cluttered figure (highest subband entropy score).

Figure 10. Figure SPM.3, panel a, showing extent of Northern Hemisphere March-April (spring) average snow cover; time-series show annual values, and where assessed, uncertainties are indicated by coloured shading. Reproduced from IPCC, 2013a.

Figure 11. Schematic of how part of a complex line (solid green line) might be decomposed into chunks by segmenting the line at points of local curvature extrema. Black circles in middle box indicate points at which the connected line may be segmented. Right-hand box shows the resulting segmentation. In this example, for simplicity, it is assumed that the area under the line is foreground and the area above the line background, to determine locations of local curvature extrema. Figure shown is Figure SPM.3 reproduced from IPCC, 2013a.

Figure 12: Presentation of experimental trial.

Figure 13: Three examples of study graphs (solid line) and associated test graphs (dashed line) shown here together. Study and test graphs both used solid lines for stimuli presentation and were shown sequentially in the experiment.

Figure 14: Line of best-fit AOI and extreme data AOI for one of the 24 study graphs.

Figure 15: Presentation of same-different and filler trials.

Figure 16: Average sensitivity (d') for amplitude different and gradient different trials in each group, with 95% confidence intervals.

Figure 17. Comparison of interaction between test graph and warning in Study 5 (left) and Study 6 (right). Error bars represent 95% confidence intervals. Note: d' values are generally greater in Study 6 than Study 5 as expected as stimuli were adapted to in Study 6 to reduce the potential for floor effects. The pattern of differences between conditions data is consistent across both experiments.

Figure 18. Average sensitivity (d') for amplitude and gradient trials by informational content in Study 6, with 95% confidence intervals. Error bars represent 95% confidence intervals.

Figure 19. Interaction between test graph and informational content in Study 6, including the no warning group. Error bars represent 95% confidence intervals.

Figure 20. Yearly global mean surface temperature anomalies (relative to 1961–1990) between 1975 and 2009. Grey shaded area indicates period of supposed pause between 1998 and 2009. Data from HadCRUT4.

Figure 21. Examples of graph stimuli, showing the three levels of recent data for a graph with a positive global trend, and an example of trend line added.

Figure 22. Example of the probe locations in relation to one of the graph stimuli. Note: only one probe appeared on any given trial in the experiment.

Figure 23. Summary of a trial.

Figure 24. Worked example of the calculation of the mean location of expectations for one participant and one cell of the study design. Each box in the 'Yes' responses column represented a single trial where a response was given; crosses in boxes indicates a 'yes' response to that trial; empty boxes indicates a 'no' response to that trial.

Figure 25. Interaction between trend line and global trend direction. Vertical dark grey bars indicate the 95% confidence interval for the mean location of expectation distributions. Vertical line at probe location 0 provided as a reference point. Light grey shaded areas indicate the full distribution of 'yes' responses for each condition.

Figure 26. Interaction between trend line and recent data. Vertical dark grey bars indicate the 95% confidence interval for the mean location of expectation distributions. Vertical line at probe location 0 provided as a reference point. Light grey shaded areas indicate the full distribution of 'yes' responses for each condition.

Figure 27. Example of ‘arrow present’ stimuli.

Figure 28. No interaction between arrow and global trend direction. Vertical dark grey bars indicate the 95% confidence interval for the mean location of expectation distributions. Vertical line at probe location 0 provided as a reference point. Light grey shaded areas indicate the full distribution of ‘yes’ responses for each condition.

Figure 29. No interaction between arrow and recent data. Vertical dark grey bars indicate the 95% confidence interval for the mean location of expectation distributions. Vertical line at probe location 0 provided as a reference point. Light grey shaded areas indicate the full distribution of ‘yes’ responses for each condition.

Figure 30. Example of vertical graph stimuli.

Figure 31. No interaction between orientation and global trend direction. Vertical dark grey bars indicate the 95% confidence interval for the mean location of expectation distributions. Vertical line at probe location 0 provided as a reference point. Light grey shaded areas indicate the full distribution of ‘yes’ responses for each condition.

Figure 32. No interaction between orientation and recent data. Vertical dark grey bars indicate the 95% confidence interval for the mean location of expectation distributions. Vertical line at probe location 0 provided as a reference point. Light grey shaded areas indicate the full distribution of ‘yes’ responses for each condition.

Acknowledgements and declarations

I would like to thank my supervisors, Prof Kenny R. Coventry, Dr Irene Lorenzoni and Dr Thomas F. Shipley for their guidance and support throughout the PhD. I would also like to thank colleagues within the Tyndall Centre for Climate Change Research, particularly Asher Minns and Prof Corinne Le Quéré, for reflective discussions, which have helped to inform my thinking. Furthermore, I am very grateful to Jackie Orford for her help in coordinating and facilitating administrative aspects related to the PhD work. In addition, I would like to thank Isla Harold and Peter Harold for their support and their willingness to act as a sounding board throughout my studies.

The candidate has not previously submitted any of this work towards the award of a degree.

Some of the work contained within this thesis has been communicated to the scientific community via publications. In all instances Jordan Harold was the primary author who drafted the text, which was then provided for secondary authors' review and input.

Parts of Chapter 1 and 5 are presented in Harold, Lorenzoni, Shipley and Coventry (2016) and Studies 4 and 5 are presented in Harold, Coventry, Lorenzoni and Shipley (2015).

Harold, J., Lorenzoni, I., Shipley, T. F. & Coventry, K. R. (2016) Cognitive and psychological science insights to improve climate change data visualization. *Nature Climate Change*, 6, 1080-1089.

Harold, J., Coventry, K. R., Lorenzoni, I., & Shipley, T. F. Making Sense of Time-Series Data: How Language Can Help Identify Long-Term Trends. In Noelle, D. C., Dale, R., Warlaumont, A. S., Yoshimi, J., Matlock, T., Jennings, C. D., & Maglio, P. P. (Eds.) (2015). *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*. Austin, TX: Cognitive Science Society; pp 872-77.

Chapter 1: Introduction

This introductory chapter presents a summary of why there is a need to support improved communication of climate science through data visualisations, followed by a review of relevant psychological and cognitive science evidence to help inform how data visualisations could be enhanced to support readers' comprehension. The chapter ends by stating the overarching goals of the research presented in this thesis and an overview of the thesis chapters.

Visualising data is integral to modern scientific practice. Scientists create visualisations to explore and analyse data and to communicate the findings of those analyses to others. In academic contexts, data visualisations for communication readily bring to mind scientifically rigorous figures published in journal articles (typically read by other scientific experts). However, scientific endeavour is also a social endeavour; research is often funded by society, is usually directly or indirectly relevant to society, and scientific findings often have the potential to change society. Critically, for science to be useful to society, scientific findings need be communicated with society – as emphasised by the UK Government Chief Scientific Advisor, Sir Mark Walport, stating that, “*Science isn't finished until it's communicated*” (Ewles, 2013, pp. 1). Scientific data visualisations have the potential to support understanding of scientific information within society, and in turn, support societal decision-making. However, data visualisations are only effective for communication if they are understood, especially if communication is to support decision-making and action.

Take climate change for example, where greenhouse gas emissions from human activities are causing the world to warm, resulting in widespread impacts to natural and human systems (IPCC, 2014a). Over recent decades, scientific research about the causes and impacts of climate change has grown rapidly (Minx et al., 2017), identifying that mitigating and/or adapting to a changing climate will require large-scale action across society (IPCC, 2014a). The Intergovernmental Panel on Climate Change (IPCC) are tasked by governments (under the framework of the United Nations) to provide policy-relevant assessments of

climate change, its impacts, and options for society to mitigate and adapt to a changing climate. Assessments by the IPCC are typically published every 5-7 years, and contain scientific data visualisations to support their communication.

However, the data visualisations of IPCC reports have been criticised for being inaccessible to non-expert audiences (acknowledged by the IPCC, see IPCC, 2016) and evidence suggests that these criticisms are valid. For example, non-expert viewers tasked with interpreting an IPCC figure of climate model projections struggled to understand it as intended by the authors of the figure (McMahon, Stauffacher, & Knutti, 2015). In this study, novice readers (academics in disciplines other than climate change, and governmental representatives) and expert readers (climate science academics) were tasked with interpreting a figure showing projected global surface warming under different scenarios through to the year 2100 (Figure 1). Novice readers typically failed to identify uncertainty related to scenarios (represented in the figure by the spread of the scenario projections – i.e. the range between the lower orange line and the upper red line), and instead attributed the uncertainty in projections to climate models. This suggests a failure of the figure to communicate an important message – namely that uncertainty in future warming is primarily due to uncertain societal choices and not due to uncertainty in climate models.

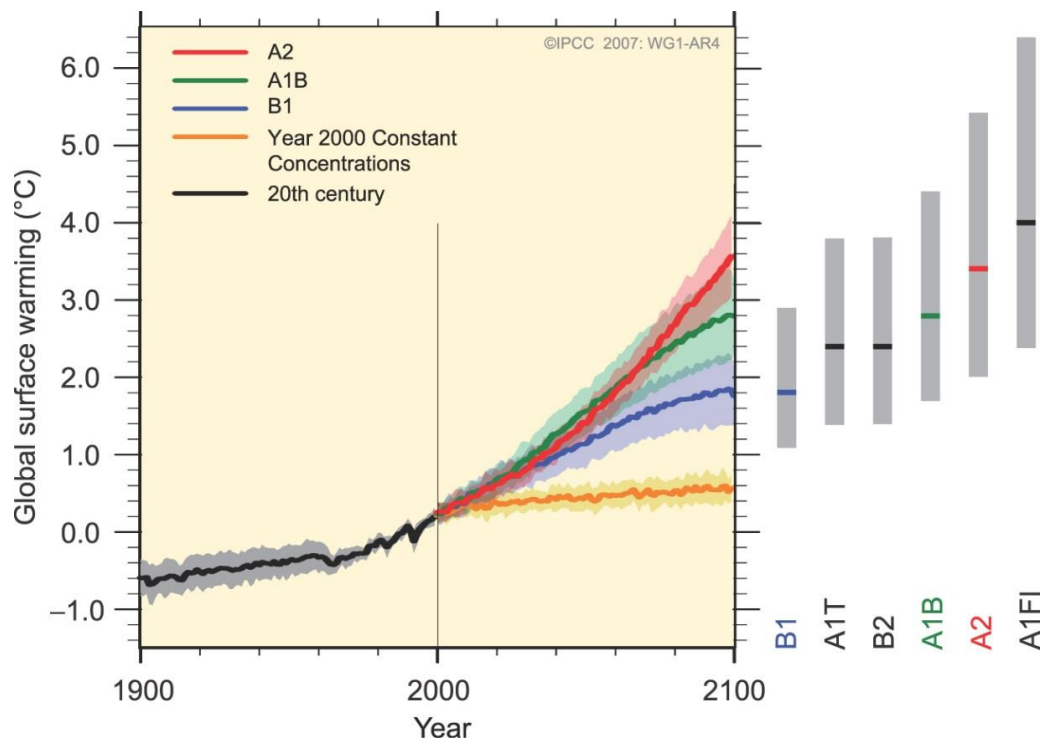


Figure 1. Figure SPM.5 from IPCC Fourth Assessment Report, Working Group 1, Summary for Policy Makers (IPCC, 2007).

Original figure caption: Solid lines are multi-model global averages of surface warming (relative to 1980–1999) for the scenarios A2, A1B and B1, shown as continuations of the 20th century simulations. Shading denotes the ± 1 standard deviation range of individual model annual averages. The orange line is for the experiment where concentrations were held constant at year 2000 values. The grey bars at right indicate the best estimate (solid line within each bar) and the likely range assessed for the six SRES marker scenarios. The assessment of the best estimate and likely ranges in the grey bars includes the AOGCMs in the left part of the figure, as well as results from a hierarchy of independent models and observational constraints.

Uncertainty is particularly challenging to visually synthesize and represent in climate knowledge, and there is a diversity in normative judgements about the implications of such uncertainties (Mahoney & Hulme, 2012). Furthermore, climate scientists and non-experts may also use different strategies to create

meaning from climate science data visualisations (Stofer & Che, 2014). In this study experts (oceanographers) and non-experts interpreted thematic maps showing oceanographic data, such as surface sea temperature, while being eye-tracked. Experts were found to make more fixations (which were on average shorter in duration) than non-experts on each visualisation, suggesting greater “meaning-making” (Stofer & Che, 2014, pp. 7). Furthermore, data visualisations of the same data represented in various visual styles have been shown to differentially influence judgements about future climate (Daron, et al., 2015). These data indicate that comprehension of scientific figures depends not only on the visual content of a visualisation, but also on parameters related to the viewer.

As outlined, there is an unmet need to support improved communication of climate science with society. One of the goals of the cognitive and psychological sciences is to understand how people comprehend written and visual information. Therefore, the evidence-base from these disciplines can offer potential solutions to support the communication of scientific data in data visualisations in contexts such as climate change. This chapter next reviews evidence from the cognitive and psychological sciences about how people construct meaning from a scientific figure and evidence relevant to enhancing the accessibility (i.e. ease of comprehension) of climate science figures, such that they can be more easily understood by non-expert audiences.

Cognition for scientific data visualisations

Data visualisations are often an effective way to communicate data - not only can they store and organise data efficiently, but they enable us to think about the data using visual perception (Hegarty, 2011). Representing data visually can create patterns that the human visual system can easily process (e.g. the iconic climate change ‘hockey-stick’ graph). However, data visualisations are not direct representations of reality; the meaning of the data they represent must be interpreted by the viewer.

Cognitive models of the comprehension of visual displays, including data visualisations such as scientific figures, posit that both the visual features of the

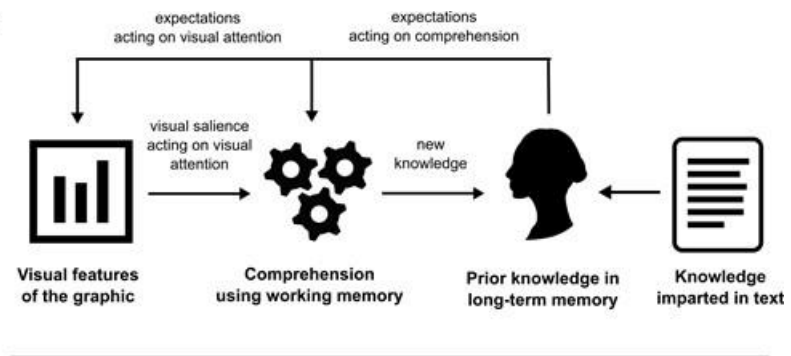
display, and an individual's prior knowledge, influence comprehension (Pinker, 1990; Freedman & Shah, 2002; Trickett & Trafton, 2006) (Figure 2). First, sensory processes direct the eyes to specific features of the display. Visual attention determines which features of the display the viewer looks at. Features that are visually salient (e.g. by virtue of their colour, shape, size) can draw the attention of the viewer – known as *bottom-up* visual processing. Conversely, the viewer's expectations, driven by prior knowledge (their previous experience of the world, and their goal or reason for looking at the display), can also direct visual attention – *top-down* visual processing (Figure 2a) (Pinker, 1990). As visual information is perceived from the features of the display, a mental representation of the information is created in memory. The nature of the mental representation is influenced by prior knowledge and goals and is constantly updated as the viewer visually explores the display (Freedman & Shah, 2002; Hegarty, 2011).

These cognitive processes are cyclical in nature; perceived and mentally represented information acts on expectations, which in turn direct further exploration of the display (Neisser, 1976). The human brain is thought to support cognition by constantly trying to match incoming sensory information against predictions of what to expect (Clark, 2013). When perceived information matches our expectations, comprehension is easy. Accessibility of a display can therefore be improved by matching visual features and prior knowledge (Figure 2b).

a. Comprehension of graphics

Comprehension involves direction of visual attention, which can be driven by the visual features of the graphic and by the viewer's expectations from their prior knowledge.

Comprehension of attended information takes place in the context of prior knowledge using working memory, and enables inferences to be drawn from the data, creating new knowledge in long-term memory.



b. Improving accessibility

Graphics can be made more accessible and more easily understood by matching graphic parameters with parameters that influence or make up prior knowledge of the viewer.

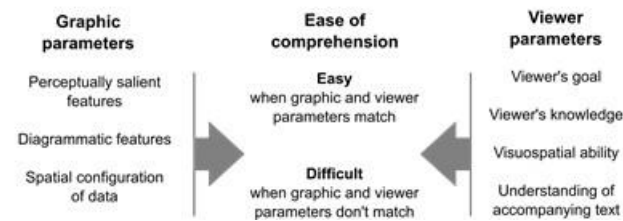


Figure 2: Conceptual overview of the process of comprehension for data visualisations and approaches to improving accessibility.

Importantly, alternative representations of a dataset that are informationally equivalent, i.e. contain the same information, are not computationally equivalent in terms of the cognitive processes involved in their comprehension (Larkin & Simon, 1987). Visualising data in graphical formats can ‘augment’ cognition, afforded by the human visual system (Hegarty, 2011; Scaife & Rogers, 1996). First, displays can meaningfully organise information in a spatial array, grouping similar aspects of the data in close spatial proximity (Larkin & Simon, 1987; Wickens & Carswell, 1995). Consequently, scientific data visualisations can provide structure to the data. Second, this visual structure can be relatively easily encoded by the human perceptual system (Scaife & Rogers, 1996). Patterns in the data can emerge when the data are visualised, such as trends in line graphs via connected lines (Hegarty, 2011). In contrast to a data visualisation, extracting such patterns from numerical or textual presentations of the data requires effortful cognitive processing (Larkin & Simon, 1987).

While existing cognitive models provide a useful framework to consider how individuals comprehend scientific figures, the nature of the cognitive processes involved in integrating perceived visual information with prior

knowledge, and the nature of mental representations in these processes, are to a large extent under-defined. For example, individuals may need to mentally animate internal representations to support inference-making and therefore it has been suggested that cognitive models should integrate a spatial processing component (Trickett & Trafton, 2006). Furthermore, cognitive insights that fall under the broad banners of ‘bottom-up visual perception’ and ‘top-down knowledge’ are rich and diverse. Identifying relevant insights from the broader cognitive and psychological literatures therefore has the potential to inform how visual parameters and viewer parameters can be better matched to support comprehension of data visualisations (Figure 2b).

Intuitions for effective data visualisations

Prior to considering the broader psychology and cognitive science literature relevant to supporting comprehension of scientific figures, it is worthwhile to briefly consider to what extent designers and viewers of data visualisations have an intuitive awareness of what makes an effective visual for communication. This is particularly relevant to scientific domains where advances in computing and software technologies have enabled scientists to create a wide-range of visual representations, as is the case in climate science (Nocke, et al., 2008). It is also important because representations may offer the viewer flexibility in how the data are displayed via interaction with the display. Such advances offer the potential to better match visual parameters to viewer parameters to improve accessibility. However, these advances also place demands on creators and viewers of data visualisations in terms of their competence in selecting effective visual representations of the data for the task at hand (diSessa, 2004).

Evidence suggests there may be limits to experts’ self-awareness (metacognition) for creating or choosing effective visual representations of data. For example, some experts, as well as non-experts, show preferences for visual features that can actually impair comprehension, such as realistic features (Smallman & St John, 2005), 3D features (Zacks, et al., 1998), and extraneous variables in data (Hegarty, et al., 2009). Consequently, intuitions about good

design practices may not always match best practice informed by cognitive principles, and viewer preferences may not always be predictive of ease of comprehension. This highlights the potential for designing data visualisations with cognitive principles in mind, and testing them with viewers. Such an approach offers an empirical approach to improving the visual communication of scientific data.

Improving accessibility of scientific data visualisations

This section reviews four key areas of psychological and cognitive science research relevant to improving the accessibility of data visualisations: directing visual attention; reducing visual complexity; supporting inference-making; and integrating text with data visualisations.

The role of visual attention

To understand the details of a data visualisation we use our central vision, afforded by the fovea centralis, which provides greater acuity than our peripheral vision. The visual field of the fovea centralis is approximately two degrees of visual angle in diameter (Rayner, 2009), meaning that when viewing an image from a distance of 60 cm (such as on a computer screen at about arm's length), our central vision covers an area approximately 2 cm wide. At any one moment in time our central vision can only focus on a limited area of a visual. Therefore, we move our eye gaze to sample information from different spatial locations (Figure 3a), and to build a detailed representation of the data visualisation as a whole we encode and retain information from these different spatial locations in memory. If visual features are not visually salient, they may not be attended to. For example, as shown in Figure 3, an individual may give little attention to the legend of data visualisation, preventing information in the legend being used to support comprehension.

Limited cognitive resources mean that only a fraction of the rich visual information entering the eyes at any given point in time is meaningfully processed

and encoded to our internal representation in memory (Simons & Chabris, 1999). Where to look, and what information to process, is directed by visual attention. Consequently, if important details in a data visualisation are not captured by our attention, they will not be processed by the brain and will not be drawn on to help comprehend and interpret the data in the visual (Figure 3b). Directing visual attention to important details can therefore make data visualisations more accessible by supporting viewers to look at aspects of the visual that afford understanding.

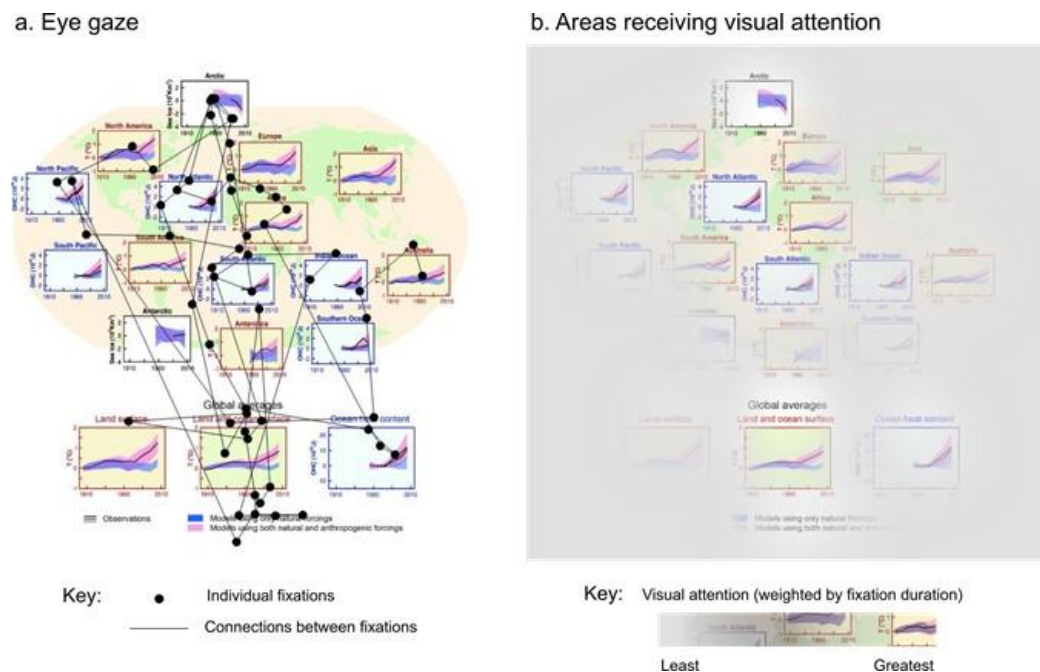


Figure 3. Example of visual attention for an IPCC figure for a non-expert viewer trying to interpret the data visualisation (measured using eye tracking: first 15 seconds of data shown). a: eye gaze shown as individual fixations and connections between fixations; b: areas receiving visual attention; computed from the locations of the fixations, weighted by the duration of each fixation.

Figure shown is IPCC, AR5, Working Group 1, Figure SPM.6 (IPCC, 2013a), original figure caption: Comparison of observed and simulated climate change based on three large-scale indicators in the atmosphere, the cryosphere and the ocean: change in continental land surface air temperatures (yellow panels), Arctic and Antarctic September sea ice extent (white panels), and upper ocean heat

content in the major ocean basins (blue panels). Global average changes are also given. Anomalies are given relative to 1880–1919 for surface temperatures, 1960–1980 for ocean heat content and 1979–1999 for sea ice. All time-series are decadal averages, plotted at the centre of the decade. For temperature panels, observations are dashed lines if the spatial coverage of areas being examined is below 50%. For ocean heat content and sea ice panels the solid line is where the coverage of data is good and higher in quality, and the dashed line is where the data coverage is only adequate, and thus, uncertainty is larger. Model results shown are Coupled Model Intercomparison Project Phase 5 (CMIP5) multi-model ensemble ranges, with shaded bands indicating the 5 to 95% confidence intervals.

Directing attention by visual design

Visual properties that can capture attention by acting on bottom-up perceptual processing include colour, motion, orientation and size (Wolfe & Horowitz, 2004). In addition, there are well-documented ‘Gestalt’ principles governing how individual elements in a visual are grouped together psychologically into meaningful entities (Bruce, Green, & Georgeson, 2003). When elements of a visual show a large degree of contrast in these properties, the contrasting visual information is automatically captured by attention and appears to ‘pop-out’ from the display (Figure 4b-4d).

Another way to direct attention is through the use of arrows. Arrows are the symbolic visual equivalent of pointing gestures, which have a widely accepted meaning of ‘look here’ and are thought to direct attention automatically (Hommel, et al., 2001). They can therefore be particularly efficient visual cues to establish joint attention between the author and the viewer for specific features in a data visualisation (Figure 4e). Of course, arrows also have other uses – such as denoting motion or temporal change – and one has to be careful not to use arrows to denote different operations within the same data visualisation.

Using these properties in the visual design of climate science figures can therefore help guide attention. Particular visual properties (or combinations of

these properties) to direct attention may be more suited than others, depending on the context in which they are used.

Informed by human behaviour and neuroscience, computational models of ‘bottom-up’ visual attention have been able to accurately predict which features of an image are most likely to be attended to (Itti & Koch, 2001). Such models provide immediate assessments of visually salient features of a visual display, and might be useful to inform the design process (Rosenholtz, Dorai, Freeman, 2011). To check viewers’ actual visual attention for a data visualisation, eye-tracking can provide empirical evidence to inform visual design. For example, eye tracking has been used to observe differences in the eye movements of individuals who were successful or unsuccessful in solving a problem scenario depicted in a visual display; visual elements that supported problem solving could then be made more visually salient (Grant & Spivey, 2003).

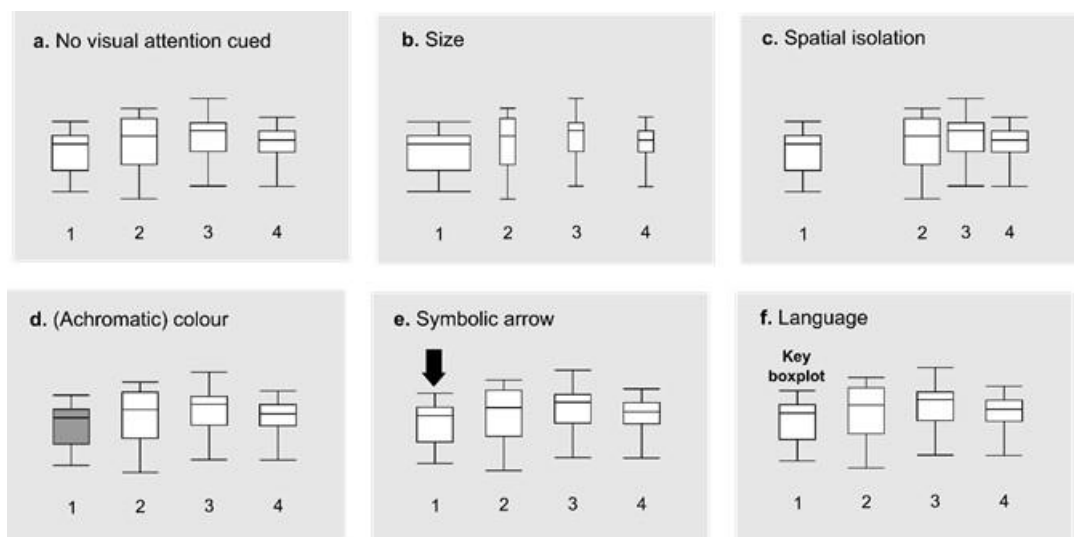


Figure 4. Schematic of properties known to direct visual attention that can be used in the design of data visualisations to help direct viewers’ attention to important information.

Directing attention by informing expectation

The details that are looked at within a data visualisation can also be directed by expectations about the task at hand. For example, patterns of eye gaze are different when viewers search a visual display for a specific feature, compared to when they try to memorise the visual display as a whole (Henderson, Weeks, & Hollingworth, 1999), or when a map is studied to learn routes as opposed to the overall layout (Brunyé & Taylor, 2009). Explicitly stating the intended task for which the data visualisation was created can help guide viewers' visual attention to appropriate information. Furthermore, prior knowledge about the data, and prior knowledge about the format or type of data visualisation chosen to represent the data, can also influence a viewer's cognition (Carpenter & Shah, 1998; Peebles & Cheng, 2003).

Research on the comprehension of meteorological charts has shown that providing viewers with relevant knowledge can support attention by directing it towards task-relevant features and away from task-irrelevant features (Hegarty, Canham, & Fabrikant, 2010). Furthermore, making task-relevant features visually salient by adapting visual design may enhance performance once appropriate knowledge is provided (Hegarty, Canham, & Fabrikant, 2010). Hence the interaction between bottom-up perceptual processing and top-down attentional control should be considered when designing data visualisations, with particular consideration given to what knowledge the viewer needs to correctly interpret the data.

Handling complexity

Some climate science figures are more visually complex than others. For example, ensemble datasets of climate models can be particularly complex and challenging to visualise (Potter, et al., 2009). What is visual complexity, and how can complexity be handled to enable data visualisations to be more accessible? Possible components that might contribute towards defining and measuring visual complexity include the number of variables and/or data points in a data visualisation (Meyer, Shinar, & Leiser, 1997), the degree of uniformity of

relationships represented by the data (Carpenter & Shah, 1998), or the degree to which the data are organised to make relevant relationships in the data easier to identify (Shah, Mayer, & Hegarty, 1999). However, while these components might be informative for simple data visualisations, they may not be easily applied across the diverse types of figures used to communicate climate science, and may not always be predictive of comprehension. For example, in some instances an increasing number of data points might make patterns in the data more obvious.

An alternative proxy for visual complexity is ‘visual clutter’, where excess visual information, or a lack of organisation of that information, impairs cognition (Rosenholtz, Li, & Nakano, 2007). Excess visual clutter can increase the time it takes to search for an item (Neider & Zelinsky, 2011), increase errors in judgments (Baldassi, Megna, & Burr, 2006), and impair processing of language accompanying a visual display (Coco & Keller, 2009). Computer models, based on principles of human cognition, can assess data visualisations for visual clutter and have been validated against viewers’ actual performance when undertaking simple tasks with data visualisations, such as searching for a specific feature (Rosenholtz, Li, & Nakano, 2007). Although such models have yet to be established as offering diagnostic value in identifying comprehension problems with data visualisations, they can be useful to inform the design process by comparing different design options for a given data visualisation (Rosenholtz, Dorai, & Freeman, 2011).

One approach to avoid unnecessary visual complexity is to only include information in a data visualisation that is absolutely needed for the intended purpose (Kosslyn, 2006). However, climate science figures may need to contain a certain level of detail or information to maintain scientific integrity (i.e. to accurately represent the extent of, or limits to, scientific knowledge). Such figures may still be visually complex in spite of only showing important information. While experts can integrate complex visual features into meaningful units of information (perceptual ‘chunks’), non-experts may lack such skills (Chase & Simon, 1973). Hence, segmenting information into chunks of appropriate size and difficulty, and guiding viewers’ attention to connections between these

components could make comprehension of the data easier (Gobet, 2005). However, such an approach should be taken with care. If the task expected of the viewer is to compare or contrast data represented in a data visualisation (known as ‘integrative tasks’), then this may be more easily performed when the data to be compared share representational similarities, such as close spatial proximity, or the same colour (Wickens & Carswell, 1995).

Supporting inference-making

Comprehension of a data visualisations of climate data goes beyond just perceptual processing of visual features. For example, enabling viewers to make relevant and scientifically robust inferences from data might be preferable to merely stating intended inferences in the accompanying text of a figure. Furthermore, data visualisations are not only used to impart information, they can also be used to support sense-making and guide decision-making. In the context of the science-policy interface, this is indeed one of the goals of science communication and aligns with the IPCC’s remit of being policy-relevant and not policy prescriptive (IPCC, 2016).

Improving accessibility to climate science data visualisations therefore involves supporting viewers to make appropriate inferences. Symbolic elements in diagrams, such as lines, boxes, crosses and circles can support inference-making about relationships in the data, based on their geometric properties (Tversky, 2005). For example, lines indicate connections, while arrows can indicate dynamic, causal or functional information (Heiser & Tversky, 2006).

Inferences may also relate to the mappings between the visual features of the data visualisation and the data that they represent. Much of our cognition of conceptual ideas is thought to be metaphorical in nature (Lakoff & Johnson, 1980). For example, *more* of something is conceptualised in mind as *up*, and so temperature is said to be *rising*; similarly, financial concepts are used metaphorically in speech with regards to limiting carbon emissions, i.e. having a carbon *budget*. Using mappings that match natural or cultural metaphors can therefore aid cognition (Lakoff & Johnson, 1980). For example, colour contains

symbolic meaning, with red usually associated with ‘warm’ and blue with ‘cold’ (Ho, et al., 2014) and indeed these colour choices are often used to represent temperature values in meteorological data visualisations. Metaphors often differ between cultures (Kövecses, 2005), and so choice of metaphors should be informed by the target audience (see section below on tailoring data visualisations to different audiences).

How data are structured in a visualisation can influence the type of information extracted, and in turn, what inferences are made about the data (Shah & Carpenter, 1995). For example, global climate projections are typically plotted as line graphs with time on the x-axis and the variable of interest (e.g. temperature anomaly) on the y-axis, which may direct viewers to consider given points in time and their associated temperature projections. Conversely, plotting temperature anomalies on the x-axis and time on the y-axis frames the data in terms of a projection of time for a given temperature threshold (Joshi, et al., 2011). Although in both cases the data are the same, the alternative graphical representations may result in viewers drawing different inferences.

Sometimes the viewer of a data visualisation may need to make inferences about the data that are not explicitly represented in the visual. Examples include making inferences about the uncertainty of the data (Trickett, et al, 2007), relationships across multiple data visualisations (Trafton, et al., 2000), and relationships between a theory and data in a visual (Trafton, Trickett, & Mintz, 2005). Such tasks involve spatial reasoning, i.e. the viewer must mentally infer information through spatial processes and transformations (Trafton, et al., 2002). In such cases, inferences can be supported either by explicitly showing the inferences in the data visualisation (and so removing the need for spatial processing), or by supporting viewers’ spatial reasoning, for example by using text accompanying the visual (see below).

Using text to support cognition

Visualisations of climate data are rarely used in isolation of accompanying text - text labels typically indicate the referents of the data, such as what the axes and

data points represent. In accordance with norms of scientific reporting, captions provide contextual information and are placed under figures, while the relevance of the figure and inferences that can be drawn from it are placed in the body text, sometimes spatially distant from the visual.

Separating text from data visualisations comes with a cognitive cost, known as the *spatial contiguity effect* (Mayer, 2009). When there is distance between the spatial locations of the text and corresponding visual, attention must be split between the two. The viewer must visually search for the corresponding elements (i.e. moving from text to visual, or vice versa) and then integrate both sources of information. Viewers may not exert effort to do this and instead may simply treat text and data visualisations as independent units of information and read them independently of one another (Holsanova, Holmberg, & Holmqvist, 2009). However, when the distance between text and visual is reduced, less searching is required, and connections can be more easily made, resulting in improved comprehension (Ginns, 2006). Tightly integrating text and data visualisations has been advocated as good design practice to support comprehension, i.e. embedding text *within* a visual (Figure 4f), or even embedding small data visualisations *within* text (Tufte, 2006).

Furthermore, language that accompanies a data visualisation has the potential not only to provide context, but also to influence thought about the spatial relationships of the properties of the visual. Tasks involving spatial relationships might include comparisons of temperature anomalies at different spatial locations on a map, inferring trends in data from observed time-series data (which spatially plot x-y relationships), or comparing uncertainty ranges for future projections of climate under different scenarios. These tasks all involve spatial cognition, i.e. thinking about spatial relationships. Attending to linguistic information while looking at visual information is known to influence spatial cognition, such as supporting spatial reasoning (Loewenstein & Gentner, 2005). Language can also influence the extent to which a static visual is mentally animated and the manner in which it is animated (Coventry, et al., 2013), which again might help with spatial reasoning. Accompanying text can therefore support viewers in making appropriate spatial inferences from a data visualisation.

Tailoring data visualisations to different audiences

So far, insights drawn from *general* principles of human cognition to help inform improved visual communication of climate science data have been considered. However, it is important to acknowledge that certain cognitive factors may differ between audience groups, and between individuals within those groups.

Colour is one area where there is marked individual and cultural variation. People who experience colour-blindness perceive colours differently from the general population and so colour choices for scientific figures should be carefully chosen to avoid perceptual difficulties (Light & Bartlein, 2004). The native language one speaks can also influence colour perception – the number of colour terms available in a language can influence colour discrimination (Thierry, et al., 2009), which might result in perceptual differences in the boundaries of colour-mapped data. Such problems can be avoided by using achromatic (e.g. greyscale) colour mappings in which data values are mapped to luminance rather than hue (Moreland, 2009), or by using colour scales that enable easy differentiation of colour (Harrower & Brewer, 2003).

As well as perceptual differences, there are also group differences in higher-level cognitive skills, such as spatial reasoning. Experts often have strong spatial reasoning skills, as has been shown in the geosciences (Shipley, et al., 2013), whereas spatial reasoning by non-experts may depend on their general visuospatial abilities (Hambrick, et al., 2012). Moreover, how attention is directed across a page exhibits marked cultural variations, with reading direction in a language (e.g. English – left to right; Arabic – right to left) associated with the direction of attention in visuospatial tasks (Shaki, Fischer, & Petrusic, 2009).

Other differences are more tied to an individual's personal knowledge and experience. For example, prior experience can lead to a knowledge of 'where to look' and so can limit visual attention to specific spatial locations (Torralba, et al., 2006). Similarly, the extent of prior knowledge about the data being visualised and prior experience using specific graphical formats can influence the ease with which inferences can be drawn from data (Ratwani & Trafton, 2008). There can be trade-offs between using an unfamiliar graphical format that may be difficult

to initially interpret but which efficiently represents a set of data, and a more familiar format whose structure can easily be grasped but which may provide an inefficient representation of the data (Peebles & Cheng, 2003). Individuals may hold different and sometimes inaccurate mental models about complex scientific systems (Gentner & Gentner, 1983), such as the underlying physical principles of climate change (Sterman & Sweeney, 2007). Understanding a viewer's existing mental model about the data and the systems from which the data originate can inform how they can best be supported to make scientifically robust inferences.

While comprehension of a data visualisation can be dependent on such factors outlined above, the underlying mechanisms responsible for human cognition are shared by everyone. Hence, general principles drawn from human cognition can inform approaches to improve the accessibility of data visualisations, but the specific way in which they are applied needs to be tailored. Consequently, testing of data visualisations is important to ensure they are comprehensible to achieve the desired communication goals (McMahon, Stauffacher, & Knutti, 2015; Hegarty, 2011).

Gaps in current knowledge

Despite advances in our understanding of the comprehension of data visualisations, there are important gaps in current knowledge that are of direct relevance to visualising climate data. Uncertainties of data can be difficult to communicate (Gigerenzer, et al., 2005; Budescu, Broomell, & Por, 2009). Although general principles have been proposed for visually communicating probabilistic uncertainty, the deep uncertainties of climate change, in which knowledge and values are often disputed and outcomes are dependent on human behaviour, may not easily translate into visual representations (Spiegelhalter, Pearson, & Short, 2011). Further research is needed on how different visual representations of uncertainty might support or hinder decision-making (Andrienko, 2010), and the cognitive processes involved in such tasks.

To provide decision-makers with access to data tailored to their needs, researchers and climate service providers are exploring the use of interactive web-

based data visualisations, such as The Climate Explorer (part of the U.S. Climate Resilience Toolkit) (toolkit.climate.gov) and The IMPACT2C web-atlas (atlas.impact2c.eu). Interaction, such as filtering or highlighting task-relevant information (Crampton, 2002) has the potential to support comprehension. However, there can be large individual differences in the degree to which people use interactive functions and the extent to which they use these functions effectively (Cohen & Hegarty, 2007); viewers require competence in meta-representational skills to make appropriate interactions (diSessa, 2004). Consequently, unless viewers have the required skills, there may be limits to how useful interactive data visualisations are to support comprehension and accessibility.

Both interactive data visualisations and animated data visualisations have been suggested to support the outreach of future IPCC assessments (IPCC, 2016). Research comparing static visuals with animated visuals is often confounded by additional information being provided in animated visuals; hence observed benefits of animation in some tasks may not be due to animation per se (Tversky, Morrison, & Betrancourt, 2002). In some cases animation may impair comprehension (Mayer, et al., 2005). Viewers may extract perceptually salient information rather than task-relevant information from animations (Lowe, 1999; Lowe, 2003) and cognitive processing of the visual information may not be able to keep up with the pace of the animation (Hegarty, Kriz, & Cate, 2003; Lowe, 1999). Animating data visualisations might be beneficial in specific situations if cognitive demands of processing the information are factored into the design of such visuals (Griffin, et al., 2006). Providing an element of user-control offers the potential to overcome some of these information processing limitations (Betrancourt, 2005). The decision to use an animated or interactive data visualisations over a static visual should be informed by cognitive demands and task requirements, be designed taking cognitive principles into account, and be tested with viewers to check comprehension (Shiple, Fabrikant, & Lautenschütz, 2013).

Together with the gaps identified above, there are also limitations in the extent to which existing cognitive models (as outlined at the start of this chapter)

reflect cognitive processes involved in the comprehension of complex real-world data visualisations. The evidence-base on which existing cognitive models are based is largely drawn from studies involving comparatively simple datasets involving comparatively simple tasks (Hegarty, 2011). Consequently, there is a risk that theoretical work does not reflect the range of cognitive processes involved in more ecologically valid contexts. This has led to calls to advance translational research between cognitive science and applied disciplines (Fisher, Green, & Arias-Hernández, 2011; Hegarty, 2011). Such an approach is gaining ground in disciplines such as cartography (Fabrikant, Hespanha, & Hegarty, 2010) and geoscience (Shiple, et al., 2013), but there remains an opportunity to do so in the context of climate science.

The purpose and outline of this thesis

The over-arching goal of this thesis is to advance understanding of cognition of scientific data visualisations (i.e. scientific figures) relevant to real-world contexts, using climate change and the work of the IPCC as an example. This is achieved through two strands of complementary work. The first strand uses mixed methods to understand the goals, contexts and constraints of the IPCC's communication of climate change via scientific figures. The second strand uses experimental methods to elucidate cognitive processes involved in the comprehension and interpretation of data visualisations, using stimuli inspired by, and analogous to, those used in real-world contexts. Here, inferences that are thought to require spatial processing are considered. Bringing the two strands together in this thesis provides two clear opportunities. First, the opportunity to draw on insights from real-world contexts to inform theoretical research on the comprehension of data visualisations. Second, the opportunity to translate research evidence from the cognitive and psychological sciences into practice to support communication of scientific knowledge within society - specifically in the context of IPCC communications.

Chapter 2 presents further context regarding the work of the IPCC and then presents evidence from interviews with IPCC authors to understand the

purpose of scientific figures in IPCC Summaries for Policy Makers (SPMs), who their intended audience is, and the context in which figures are communicated. This chapter also compares the views of experts and non-experts on the ease of comprehension of ten figures from an IPCC report to explore whether there may be differences between these groups.

Chapter 3 then investigates cognition for time-series graphs that exhibit both underlying trends and short term-variability (i.e. noise) in the data – analogous to graphs used in climate science to communicate patterns in climate data. Furthermore, this chapter also considers to what extent language (verbal instruction) might support cognition for inferring trends in noisy data.

Chapter 4 follows-on from the work in Chapter 3 by asking whether particular aspects of the data when plotted in time-series line graphs influences individuals' expectations of how the data will evolve into the future. Here perceptual design features (trend lines, directional arrows, and the orientation of the graph), are also investigated to see to what extent they influence expectations for future data.

Chapter 5 discusses findings across the research studies and reflects on how the findings can help inform further research on cognition of scientific data visualisations. Research evidence is synthesised into a set of cognitively inspired guidelines to support producers of scientific figures to enhance the accessibility of their data visualisations while maintaining the scientific integrity of their content.

Chapter 2: Are there comprehension difficulties with IPCC figures?

This chapter outlines the role of the IPCC and provides a brief review of the communication challenges associated with the outputs from the IPCC. Three associated studies are then presented, which together provide insights on how climate change data visualisations produced by the IPCC are communicated at the science-policy interface and potential challenges for their cognition.

Study 1 presents a thematic analysis of interviews with IPCC authors, identifying that a significant constraint when producing data visualisations for policy audiences is to ensure that scientific accuracy is upheld. Furthermore, the interviews highlight that data visualisations in the report are not designed for policy-makers or non-experts, but rather they are designed for other experts. Non-experts are expected to need the support of experts to understand the data visualisations.

Study 2 reports on a set of sort tasks involving the ten figures from the IPCC Working Group 1 Summary for Policy Makers (SPM) with the same group of experts in the interviews and with a group of non-experts (university undergraduates). The results indicate that experts have a good appreciation of the types of figures that non-experts feel are more difficult to understand (relative to other figures in the report) and that some of the most policy-relevant figures are expected to be particularly difficult for non-experts to understand.

Study 3 then investigates whether non-experts' perceptions of the ease of comprehension of the ten figures is associated with the visual complexity of the figures, as measured via a computational measure of visual clutter. Findings suggest that greater visual complexity is positively associated with greater perceived comprehension difficulties.

The IPCC have stated a desire to communicate outputs of future reports and assessments such that they can be understood by non-expert audiences (IPCC, 2016). The findings across these three studies highlight a need to develop

approaches to creating scientifically rigorous data visualisations that are more accessible to non-experts than those currently available.

The IPCC - background context

Reports by the IPCC provide robust scientific assessments of current knowledge regarding climate change, related to: the physical science basis (Working Group 1) (IPCC, 2013a); impacts and adaptation (Working Group 2) (IPCC, 2014b); and mitigation (Working Group 3) (IPCC, 2014c). Each working group produces their own report and, in addition, the IPCC produces a synthesis report summarising and integrating findings from across the three working groups (IPCC, 2014a). Each group report consists of underlying chapters which provide detailed assessments, and a Summary for Policy Makers (SPM), designed to highlight key information to support governments in decision-making.

IPCC assessments are typically conducted every 5-7 years, with the most recent being the fifth assessment report (known as AR5), published in 2013-14 (IPCC 2014a). Reports are written by author teams, undergo extensive peer review and are formally accepted by 195 national governments (IPCC, 2013b). As such, IPCC reports are held in high regard within the scientific community and by the national governments – the IPCC was a joint recipient of the Nobel Peace Prize in 2007 (IPCC, 2012) and the IPCC receives continued support via the United Nations (IPCC, 2013b) with the sixth assessment report commissioned for publication in 2021-22 (IPCC, 2017a). Furthermore, the release of IPCC reports attracts wide media coverage (Boykoff & Boykoff, 2007; Barkemeyer, et al., 2016) and commentary on social media (O’Neill et al., 2015; Newman, 2016), highlighting a high degree of societal interest in the work of the IPCC.

Communication challenges with IPCC reports

IPCC reports have been criticised for being inaccessible to many non-experts, with a particular focus on the complexity of the language used in the SPMs (Barkemeyer, et al., 2016; Hollin & Pearce, 2015; Budescu, et al., 2014). The

figures within SPMs (that is, data visualisations of scientific information in the form of graphs, diagrams, thematic maps and other visuals) may also be inaccessible to non-experts. For example, viewers looking at figures of climate model projections can confuse scenario uncertainty (that is, unknown future societal choices) with model uncertainty (McMahon, Stauffacher, & Knutti, 2015). There are also challenges in visually synthesizing and representing uncertainty in climate knowledge, and diversity in normative judgements about the implications of such uncertainties (Mahoney & Hulme, 2012). Accordingly, IPCC SPMs have been critiqued as being “summaries for wonks” (Black, 2015, pp. 282) – in other words, documents that can only be understood by experts, and which are not fit for purpose for use by policy makers, such as government ministers, or by the general public.

The IPCC is aware of the need to make information more accessible to broader audiences in society (IPCC, 2016). IPCC authors have responded directly to these criticisms, highlighting how communication of AR5 has changed since the publication of AR4 in 2007. For example, by providing headline statements for the Working Group 1 report, which consists of a concise (2-page) summary of the report (Stocker & Plattner, 2016); delivering press conferences with the media to explain key findings and answer questions (Jacobs, et al., 2015) and widening the provision of outreach activities, such as presenting key IPCC findings at various events (IPCC, 2016). In addition, organisations other than the IPCC create what are known as ‘derivative products’ (IPCC, 2016). These are communications that are adapted from IPCC materials for specific purposes and audiences – for example briefings for business (Symon, 2013). However, despite these efforts there still appears to be a disconnect between a demand for more accessible communication outputs and the IPCC’s supply of highly technical communications (IPCC, 2016).

There are strong arguments for the IPCC to make reports more accessible to broader audiences beyond just experts. First, climate change is a societal issue that has profound implications across the world including on energy production, food security, biodiversity and health (IPCC, 2014a). Therefore, the work of the IPCC is highly relevant to broader society. Second, the work of the IPCC is

funded by national governments, via the United Nations (IPCC, 2013b), and so tax-payer money is spent enabling the IPCC to conduct their work. Therefore, broader society has a vested interest in IPCC outputs. Third, there is clearly an appetite from broader society to access information about climate change. Climate change issues regularly attract media coverage (Schmidt, Ivanova, & Schäfer, 2013) and there is growing demand for tailored climate information for decision-makers, for example within industry and local government (Vaughan & Dessai, 2014).

IPCC data visualisations

Data visualisation, in the form of scientific figures, is an integral component of the IPCC reports. Each SPM contains 8-14 figures capturing specific aspects of each working group's assessment (IPCC, 2013a; IPCC 2014a; IPCC 2014b; IPCC 2014c). These figures, or a selection thereof, are typically presented at press conferences at the launch of the reports and are also re-used in slide kits provided by the IPCC (IPCC, 2016).

On the face of it, these figures appear to be created for policy makers, as per the name - 'Summary for Policy Makers' – would suggest. However, 'policy-makers' is a broad term that can encompass junior civil servants right through to senior politicians, who may work in wide range of diverse agencies, and whose role may vary; for example, they may be a decision-maker or an advisor (Tyler, 2013). The IPCC SPMs do not state specifically who the reports are aimed at, how they are intended to be used, or what prior knowledge is needed to understand and interpret them. Here it is important to note that national governments review and approve the SPMs, and so it may be the case that the IPCC and the national governments have an implicit knowledge of who the SPMs are created for, even if this isn't explicitly stated.

One might assume that IPCC authors make active decisions and choices with regards to how climate science data and evidence are visually represented in the IPCC SPM figures. There may for example be norms and constraints that

influence design and communication choices. Therefore, to constructively critique the IPCC figures as communication devices and identify how they might be made easier to understand, it is important to first establish who they are intended to be used by, how they are intended to be used, and the main factors that influence how they are produced. Understanding these aspects will provide important context to the communication challenges when using data visualisation of climate science. This was the purpose of the three studies outlined in the next sections below.

Study 1: What factors influence the production and communication of IPCC figures?

This study set out to understand the context in which the IPCC AR5 Working Group 1 SPM figures were produced. Qualitative research interviews with IPCC authors were conducted to identify the main factors that influence the production and communication of the figures for the IPCC AR5 Summary for Policy Makers (SPM). To achieve this goal, a series of sub-questions were developed: who are the specific audiences that the SPM figures are created for?; what is the process for the creation of SPM figures?; what are the criteria for the inclusion of a figure in the SPM?; how are the SPM figures communicated?; which figures are difficult for audiences to understand?; and which figures are most important for future climate policy?

Method

Interview questions were designed to explore the specific topics as listed above in the research aims. Interviews also collected data regarding interviewees' areas of expertise, and role in the AR5 report. Interviews were conducted either in person, or remotely via Skype. In addition to open-ended questions providing qualitative data, sort-tasks (with the ten AR5 Working Group 1 SPM figures) were used to collect quantitative data on perceptions of ease of comprehension and importance to inform future climate policy of the figures. Interviews were conducted between March 2014 and February 2015.

Participants

A total of 18 interviews were conducted. Seventeen were with individuals listed either as Drafting Authors or Draft Contributing Authors to the IPCC AR5 Working Group 1 Summary for Policy Makers (IPCC, 2013a) and one interview was conducted with an author who had worked previously with the IPCC. There was an 82% participation rate (22 authors contacted to take part, 18 agreed).

The interviewees were all established climate change research scientists, employed by universities or research institutes. IPCC authors are selected to contribute to the reports based on their high-level expertise, and they volunteer their time to work on IPCC reports (IPCC, 2013b).

Stratified sampling was employed to ensure that, where possible, variation in geographic representation, gender, role in authorship, and area of expertise (across the report's underlying chapters) was reflective of the full set of authors to the AR5 WG1 SPM (Table 1). Across all interviewees, there was representation from 11 of the 14 chapters of the main IPCC AR5 WG1 report. A full breakdown of interviewees' areas of expertise across chapters is not presented here to avoid the potential for breaching interviewees' anonymity.

Table 1. Study 1 comparison of demographics of the interviewed sample to the full set of authors to the AR5 WG1 SPM.

	All listed authors (n=71)	Interviewed sample (n=18)
Continent		
Europe	39 (52%)	9 (50%)
North America	22 (29.3%)	5 (27.8%)
Australasia	7 (9.3%)	2 (11.1%)
Asia	4 (5.3%)	1 (5.6%)
South America	3 (4%)	1 (5.6%)
Middle East	0 (0%)	0 (0%)
Africa	0 (0%)	0 (0%)
Gender		
Male	60 (84.5%)	15 (83.3%)
Female	11 (15.5%)	3 (16.7%)
Author role		
Drafting author	34 (47.9%)	8 (44.4%)
Draft contributing author	37 (52.1%)	10 (55.6%)

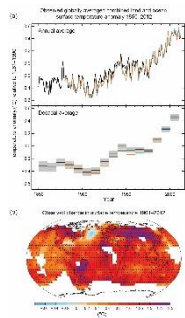
Participant recruitment

Participants were initially emailed an invitation to take part in the interview via an intermediary (who was a contributor to the IPCC AR5 report and professional colleague of the authors). If individuals expressed an interest in taking part, further details about the study, and a consent form were then emailed by the researcher and a date scheduled for the interview.

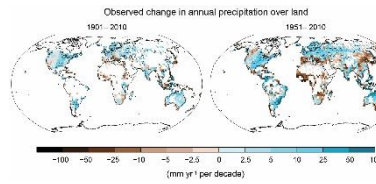
Semi-structured interview protocol

The semi-structured interview protocol was developed to ensure consistency across interviews, and enable flexibility in response to topics raised by the interviewees. The interview questions covered the following topics: the audiences of the figures; the purpose of the figures; the process through which the figures are created; strengths and weaknesses of the figures; and the use of the figures by the IPCC and by others. (See Appendix 1 for the full interview protocol).

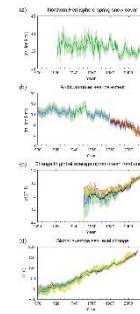
Ten A5 cards containing the ten AR5 Working Group 1 SPM figures, without their associated captions, were provided to participants at certain stages of the interview to aid their thinking (Figure 5).



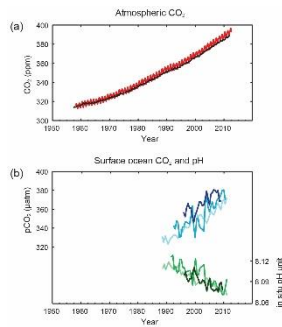
1. Figure SPM.1



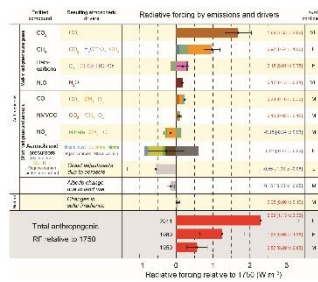
2. Figure SPM.2



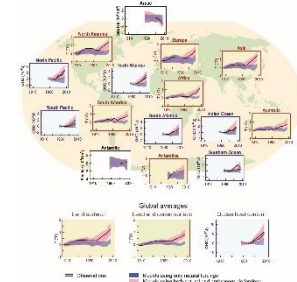
3. Figure SPM.3



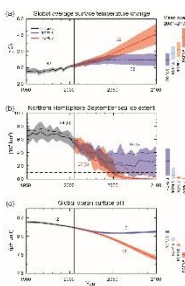
4. Figure SPM.4



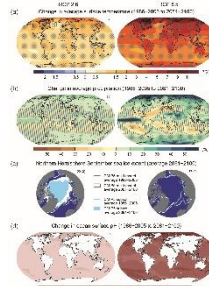
5. Figure SPM.5



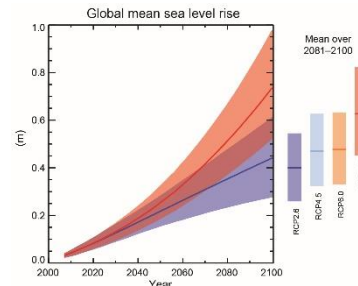
6. Figure SPM.6



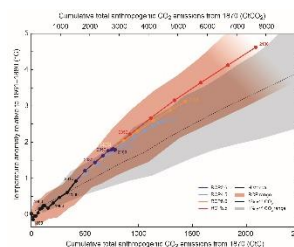
7. Figure SPM.7



8. Figure SPM.8



9. Figure SPM.9



10. Figure SPM.10

Figure 5. Thumbnails images of the ten figures from the IPCC AR5 Working Group 1 Summary for Policy Makers (IPCC, 2013a). Larger versions of figures are not provided here due to copyright, but can be accessed in reference IPCC, 2013a.

Sort-tasks

Three sort-tasks, using the ten figures from the IPCC AR5 WG1 SPM as stimuli, were interleaved with the interview questions. At the start of the interview, participants were asked to “*Rank order the Figures from the one you think university undergraduates without climate science training would find easiest to understand through to the one that you think they would find the most difficult to understand*”. A second sort-task mid-way through the interview asked participants to rank the figures on ease of understanding by policy-makers. A final sort-task at the end of the interview asked participants to “*Rank order the Figures based on their importance to help inform future climate policy, from the one you think is the most important through to the one that you think is least important.*” Further details of these sort-tasks and analyses and results are presented in Study 2 and are therefore not mentioned further in in this section.

Procedure

Interviews were conducted via video-conference or face-to-face and in both cases lasted approximately 1 hour. For participants taking part remotely via Skype, an interview pack was mailed to them providing the same set of materials as those used with participants who were interviewed in person. After initial introductions, key points from the information sheet were described to the participants and there was an opportunity for any questions or clarifications. Participants then gave informed consent prior to the start of the interview. All interviews were audio-recorded (with the consent of participants) to enable verbatim transcription, with recording starting at the first interview question.

The interview protocol was followed to guide the overall structure of the interview. Additional follow-up questions and clarification questions were asked by the interviewer in order to explore answers in more depth. At the end of the interview, participants were offered the opportunity to add any additional comments or clarifications. Participants were then debriefed.

Analysis and results

Analytical approach

To identify the predominant patterns in the qualitative data across the interviews, thematic analysis was used to extract themes that identify the main factors that influenced the production and communication of the figures for the IPCC AR5 SPM. Thematic analysis is a flexible tool to code qualitative data using a rigorous and systematic approach, while acknowledging the ‘active’ role that the researcher takes in conducting the analysis – i.e. prescribing meaning in a given context (Braun & Clarke, 2006).

A ‘critical realist’ approach was adopted (Clarke, Braun, & Hayfield, 2015), in which the data analysis reports the production and communication of the SPM figures from the perspective of the authors, but acknowledges that these experiences are formed within the broader context of the use of science and its communication in society. Further, the analytical approach was inductive, whereby themes were identified by keeping as close as possible to the semantic meanings within the data. Descriptive reporting is accordingly adopted to summarize and describe the identified themes (Clarke, Braun, & Hayfield, 2015).

The thematic analysis was conducted as per the six phases outlined by Braun and Clarke (2006). Interviews were transcribed verbatim, covering the full length of the interview, except for initial introductions. Initial codes were generated to identify interesting aspects in the data. Codes were then mapped to identify similarities, links between codes and to search for potential themes. Candidate themes were then reviewed back to the data, at which point some themes were dropped if they lacked adequate support across interviews. Themes were then refined, defined and named. Data extracts were then selected to illustrate the themes. Finally, the analysis was contextualised back to the research question.

Thematic analysis

Three main themes were identified from the analysis that provide insight to the factors affecting the communication of climate science in the Figures of the IPCC AR5 WG1 report, which were: ‘scientific rigour’, ‘useful science for experts’, and ‘inaccessible without expert guidance’. Together, these themes outline a three-point argument:

- 1) Due to the IPCC’s remit to produce a scientifically robust report of the current knowledge on climate change there is a perceived limit to which information can be simplified without losing accuracy. The information presented therefore retains complexity (**‘scientific rigour’**).
- 2) Although information is selected and structured for its relevance to policy-makers, the complex information is actually aimed at government experts and the scientific community (**‘useful science for experts’**).
- 3) Consequently, many of the figures of the report are not expected to be understood by non-experts unless they receive additional support and explanation from experts (**‘inaccessible without expert guidance’**).

Within each theme, sub-themes were identified that provided further nuance and context to the main themes. A summary of the main and sub-themes and their definitions are provided in Table 2. In the following sections, each theme is defined in detail and quotes extracted from the interviews are presented to demonstrate evidence for each theme/sub-theme.

Table 2. Study 1 definitions of themes and their sub-themes

Theme / sub-theme	Definition
1. Scientific rigour	Requirement for the SPM to be a scientifically accurate document.
1.1 Chapters as source	Information presented in the SPM must have a 'line of sight' back to the report chapters.
1.2 Experts review and amend	Climate change experts are responsible for evaluating and editing figures.
1.3 Complexities retained	Detailed scientific aspects of the information are kept in the figures.
2 Useful science for experts	The SPM is created with the aim of providing a functional document that meets the needs of expert readers.
2.1 Policy relevant	The primary purpose of the SPM is to communicate policy relevant information needed for decision-making.
2.2 Story-telling	Figures are used to highlight key messages, which together make up a narrative.
2.3 For technical analysts	The SPM figures are produced not for policy-makers per se, but for experts that work within governments.
2.4 For the scientific community	The SPM figures provide a useful resource for the scientists – both those involved in the IPCC process and scientists who are not.

Table 2 (continued).

Theme / sub-theme (continued)	Definition (continued)
3. Inaccessible without expert guidance	Non-experts are expected to find the figures difficult to understand unless they receive support from experts.
3.1 Visual complexity	Informational complexity results in some figures that contain a lot of content and so are difficult to understand.
3.2 Visual formats	Certain graph types that are considered ‘familiar’ are thought to be easier for people to understand than less familiar graph types.
3.3 Expert explanations	Experts have to explain the information presented in figures in order for non-experts to understand them.

Scientific rigour (1)

This theme relates to the emphasis placed by interviewees on the report being a scientifically accurate document in line with the IPCC’s remit of providing ‘rigorous and balanced scientific information’ (IPCC, 2013b, pp. 1).

Consequently, the retention of complexity, such as the inclusion of uncertainties and the use of multiple datasets is seen to be necessary, (to highlight key elements of quotes, emphasis has been added in bold text):

*“The role of the IPCC is to produce documentation with uncertainties of what’s going on and what’s projected, and so it cannot be watered down. **All of the information that is conveyed has to be accurate** and it has to be pretty much sufficient to get across the full picture. So, it’s not journalism.*

It's not a text book either. It's a report. So, it needs to be aimed at probably the highest technical level.” (05)

*“... there is **a desire to have the technical content be pretty accurate**, and maybe in some sense complete, and so the figure can be get a little bit complicated in consequence of that I guess.” (11)*

*“...it's a scientific assessment – so that makes that **the outcome will always be scientific** and not kind of popularizable.” (14)*

Scientific rigour is often framed in the context of having to be able to defend the report in the face of potential criticism. The authors are aware that the publication of the IPCC reports has a high profile and that errors in past reports has led to criticism of the IPCC:

*“So the summary needs to be a scientific document that is completely **waterproof** in every aspect.” (18)*

*“Well I think all the figures, erm, fulfil the requirement of our requirement to be scientifically sound, and scientifically robust and **defendable...**” (17)*

Within this theme, three sub-themes were identified that illustrate how the requirement for scientific rigour influences the content of the SPM figures, which are now summarised. Because there is a need to maintain scientific rigour, figures in the SPM are brought up from underlying chapters of the report (**‘Chapters as source’**) and experts are responsible for reviewing and revising the content (**‘Experts review and amend’**). As a consequence of these processes, the content of the SPM figures retain a high level of complexity (**‘Complexities retained’**).

Chapters as source (1.1)

This sub-theme identifies that to maintain scientific rigour, information presented in the SPM must have a ‘line of sight’ back to the underlying chapters:

*“I think we were scientifically sound and robust, that all of these messages that are in these figures are kind of traceable to the text in the SPM, are **traceable to the underlying chapters** at the level of the executive summaries. So, there is nothing that pops up suddenly at the level of the SPM figure.” (17)*

*“Well I think its.. in general it's a pretty robust procedure and that really **the figures are all directly traceable to the main body of the report** and that's where all of the agonizing over the content of the figures and how to present them.” (01)*

In terms of the SPM figures, this places restrictions on the extent to which the information can be tailored for presentation in the summary document. While the figures are not simply transposed from the chapters to the summary, interviewees felt that there must be a certain degree of consistency between chapters and the summary in not only the informational content, but also to a certain extent in the visual presentation of that content:

“I'm not quite sure how far we are actually allowed to stray away from what's in one of the underlying chapters. I presume we can to some extent, but that may be a significant restriction.” (15)

“So the data that's in those figures has to appear in the chapter, but I'm not sure if the figures look exactly the same. So the figures in the SPM are not the same – the facts are the same.” (11)

“All of them come from chapter figures and then may have gotten massaged in the SPM to become more visually accessible.” (05)

Experts review and amend (1.2)

Authors also attributed the scientific rigour achieved by the report to the comprehensive expert review and editing of the content, including the figures. These review processes were typically framed as being quality control checks to ensure accuracy of content, but reviewing for relevance and comprehension/communication of the information was also mentioned. Hence, the authors expressed an awareness that the information is challenging to communicate and that the review process provides an opportunity to improve the communication of the figures:

*“In our case for SPM.10 the strength is **that it went through a lot of scientific questioning and rigorous checks** that what we show there is a good representation of what we know.” (14)*

*“And then you try to narrow down and say what's the most interesting and the most relevant, and the most robust; that's also a big discussion of what are the things that we actually believe are scientifically most defensible. And then you iterate on those. And then of course **this gets into review and then you get hundreds of comments** saying ‘this is completely unreadable’, or ‘you need to do this’, ‘you need to add this’, and then there's always compromise.” (18)*

*“But the data coverage is rather different, because some datasets infill missing data and some don't infill missing data and so the decision was involved in there, at least that I contributed into the discussion of, I didn't actually make the decision, but **I threw in my informal review comments.**” (04)*

*“Basically, what we did for figures was they took all the figures that had been suggested and they projected them and then **we had a pile on and we commented on them**, and commented about what colour schemes and whether it really conveys the message.” (05)*

Efforts to improve the figures in response to review comments were described as collaborative efforts among groups of authors, in which figures were amended in response to suggestions, and then re-evaluated. Interestingly, communication and data visualisation experts were not described as being part of this process – instead authors took on all the responsibility of refining the design and layout of the figures for presentation in the SPM:

“And then we iterated over at least 50 versions of that figure over months to try to see how can we make it simpler, what things to include, what things to leave out.” (18)

“ ... we actually tried to imagine some different figures and we had a workshop in one of the lead author meetings, er, we would have spent an hour and a half probably chatting about this figure because it is a synthesis figure. Trying to work out how to present the information differently and how to kind of make it easier, but in the end, we kind of retreated back to the figure that was derived, so this is a derived figure, even though I might have designed it, it’s a derived figure from the earlier assessment report.” (09)

“So the actual design of the figure was only with authors [of the Chapter] and also then with the TSU and the Co-Chair – so basically a technical expert group of people developing that figure – and yes we would test it, but again I would test it to people that do not that have an advanced knowledge of science. Not necessarily on this topic – but for example, we tested it on a regular basis during team meetings – just put the figure up and say “OK what do you think?”, and then you explain and then you get feedback and say “Oh no that’s not” and then you just go back and work on it again. (14)

“... there is no expert on perception or visual communication or in fact any communication person directly involved. There may be people who read it and give comments but the drafting team is a science team. It’s not... it doesn’t have anyone that is coming from the communications side,

which is probably wrong. Erm, because maybe some things could be improved by involving those people from an earlier point.” (18)

Complexities retained (1.3)

Authors made associations between retaining scientific rigour and the resulting complexity of information presented in the SPM figures, with a perception that reducing complexity would come with a cost of losing scientific rigour.

Accordingly, authors were aware that some of the content in the figures was complicated to understand, but justified this based on the IPCC remit to provide a robust assessment of the science:

*“What we always felt was that we were trying to give the right figures that gave the right sort of information. Whether it be a bit complicated or not. So we weren’t, so **we weren’t looking to simplify**, we were really looking to communicate the full range of analyses of results in an impactful way.” (09)*

*“**The best we could do maintaining scientific rigour trying to simplify it as much as we can**, but... in the cold light of day we said how much of this meets the needs of government ministers, I would probably say it doesn’t meet them very well because it’s a bit too complicated - **the images are too complicated.**” (15)*

*“So perhaps we haven’t been as innovative as we would have loved to have been, but, you know, it is a solid figure – it’s got many more panels on it, far more inclusive of the ocean, sea ice, so it brings in more elements. And **that increases the complexity to some extent.**” (09)*

*“Now this is an easy one maybe, but there’s the uncertainty bars so **I think this adds some complication.**” (17)*

Useful science for experts (2)

This theme identifies that IPCC authors have a desire to make the SPM report, including figures, a functional and helpful document that meets certain needs of expert readers. Four sub-themes were elicited that characterise how information is made useful and for which types of experts. Given that the IPCC has a remit to provide information that is policy relevant (**'policy relevant'** theme), content is structured to highlight key messages relevant for policy which together make up a narrative (**'story telling'**):

*“... according to the mandate for the IPCC, **this is extremely policy relevant**, it's leading into the climate negotiations because its conveying the message that 2 degrees is a very difficult target, and at the same time also conveying information about how much emissions we can allow for a given temperature increase.” (07)*

*“**The figures that are chosen they are capable of conveying a story that these are observed changes, this is how we attribute these changes, and these are the projections, and these are the options for the future. And that thinking has been clear to me.**” (07)*

However, given the requirement for scientific rigour resulting in complexities being retained, the information within the report is not targeted to policy-makers per se, and is instead produced for experts working within governments, often referred to as 'technical analysts' (**'technical analysts'**). In addition, given the scientific rigour of the reports and their coverage across a wide range of scientific knowledges relevant to climate change, the reports are also seen to be a product created for the scientific community (**'scientific community'**):

*“It [referring to a figure] may not be very easily kind of used by the policy makers but, you know, **people who are assisting the policy makers can get a lot out of it.**” (06)*

“They’re useful to scientists and I think they’re useful to people who take an interest in climate research who have sufficient knowledge and or scientific background to interpret them. And I think that they’re of use to them – they give a nice overview of where science currently stands on these issues, so yeah.” (04)

Policy relevant (2.1)

Authors emphasised that the primary purpose of the IPCC SPM and its figures was to communicate policy relevant information to help policy makers in their decision-making. Consequently, authors expressed that the SPM figures were created with this in mind:

*“... the intended function which is **informing policy makers** of both the state of the science so that policy decisions around climate adaptation, mitigation, and so on are founded on well considered scientific evidence.” (01)*

*“... the underlying decision making about what went in, in terms of figures at least, was you know, **how important is it for policy makers to know this information** or to see this evidence, if they’re making decisions about climate change, as opposed to would they find it interesting. (04)*

Furthermore, the inclusion of a figure in the SPM was described as being a way to highlight the policy-relevance of information. For example, Figure SPM.10 (which is referred to in the quotes below) was indicated as being the most policy relevant figure in the SPM because it provides a decision-making tool enabling policy-makers to see how different levels of cumulative CO₂ emissions will affect global average surface temperature. Consequently, by the time the figures have been distilled from the underlying chapters, they are perceived as not only being robust representations of the science, but also relevant and useful:

*“So **this one is very policy relevant** in thinking about the target for what you can emit in carbon.” ... “...So this is very policy relevant.” (09)*

*“**It is probably the most policy relevant, policy making relevant figure** in there because, well in terms of mitigation policy at least, as its directly linking the available CO₂ emissions and the probability of staying below on the projections of global temperature.” (04)*

*“And also **to explain key issues that we anticipate that some policy makers may need to be aware of.**”*

– *“Can you give an example?”*

“Yes, so the A8 one which has a connection between temperature change and cumulative total CO₂ emission, so it’s an important connection, that was one of the reasons why that was highlighted.” (15)

*“The reason why I think that it is the most important is that it provides the most direct message in terms of **the main policy relevant issue** which is emissions and their connection to climate change and how much you can emit if you want to keep climate below some target like the UNFCCC aspirational target.” (01)*

Story-telling (2.2)

In addition to providing policy-relevant information, the figures were seen to be a mechanism to highlight the ‘key messages’ within the report. In other words, the figures are used to provide emphasis on what was considered important in the document:

*“... need to come up with a few figures that **summarize some of the key statements** ... ” (18)*

*“It’s one figure that we thought a lot about **how to convey that message**, rather than just presenting it in words.” (17)*

*“ ... the fact that the figure exists puts **emphasis on the importance of the information.**” (08)*

*“ ... when we were synthesising the Summary for Policy Makers, we were picking out figures for modifications, figures that had appeared in the main text that we thought **complemented the message** that we were trying to show.” (09)*

*“And in the end what you have is a very condensed, very dense, still very technical document because it is that distilled essence of a thousand paged long technical document. And that is the way it is with the figures too – they are distilled versions of what a collection of people viewed as being **the most important scientific messages to convey.**” (01)*

Furthermore, as a set, the figures and their associated messages were seen to ‘tell a story’ that communicated the contribution of Working Group 1 to the overall IPCC assessment. The story assigned to the Working Group 1 figures was typically described in three parts – first, observations demonstrate that the climate is changing, second, these changes can be attributed to anthropogenic causes, and third, projections inform us of future climate under different scenarios to support decision-making:

*“So in a sense having all these figures **providing kind of a story** from observed to climate change, to an understanding of climate change, to projections of climate change, it kind of takes into account that we also talked to the public, kind of that we have kind of a story telling that you might not need for experts at the level of the UNFCCC.” (17)*

*“And particularly for Working Group 1, which is the physical science basis for climate change. **To communicate the story.**” (06)*

*“ ... so some of them **make a story**, so sort of have to show them all, the observation of warming, the attribution, the causality, that sort of stuff, you kind of need to show them all to make a story.” (11)*

*“And I think the figures, as they are in front of us, they support, they are pillars of the entire **narrative**. It’s like signposts, orientation aids, along that way where you come from the observations, think about the causes of*

climate change and then consider the different possibilities of the different futures you have.” (02)

For technical analysts (2.3)

When asked to describe the audience for which the figures are created for, the authors described not a policy-maker audience per se (and which might encompass a wide range of different types of people in different roles), but rather a specific audience, typically referred to as ‘technical analysts’:

*“From my point of view the figures in the SPM it says for policy makers, but **actually I think it’s for policy analysts** within government departments.” (14)*

*“So it’s not necessarily, as I understand, the ministers who the figures are directly aimed at. I would say **the figures are aimed at the technical advisers** to ministers.” (12)*

*“I think it’s most likely that they would be useful for **people who are working as a policy expert** on this topic in a government somewhere.” (11)*

The role of these ‘technical analysts’ was described as being to ‘translate’ the information presented in the SPM to policy-makers. Therefore, despite the SPM stating in its title that it is for policy makers, authors had an expectation that an additional level of interpretation would be added on top of the report by the technical analysts in order for policy makers to actually make sense of, and use the information presented in the SPM:

*“I think that in my mind at least that the target audience is... so perhaps it is a bit of a misnomer that it’s a summary for policy makers. To me, **it’s targeted more at the staff of policy makers**. So typically you know if you are a minister, the minister for environment for example. You have a staff of people, some of whom have some background. That minister is not*

*necessarily very scientifically literate. I mean, often politicians come from you know a background in law or something, so they are smart people, but they are not scientists necessarily. So, I kind of regard the SPM as being directed more at the **staff that supports policy makers** as opposed to a policy maker per se.” (01)*

*“Well that’s the Summary for Policy Makers, but my impression is that policy makers will have scientific advisors or advisors, not necessarily scientific ones, so I expect that in many cases the **policy makers will ask their advisors to do the interpretation** and tell them as policy makers should take in from the report.” (04)*

*“But policy makers, they are not stand-alone readers. So each policy makers will have, er policy maker will have a kind of **battery of people who are well-versed with science to assist them**. So to that extent, these figures, at least seven figures out of ten probably could be understood by the policy makers to a certain extent. But I still feel last two or three figures it would be very difficult to them.” (06)*

Furthermore, authors emphasised that the ‘technical analysts’ were considered to be intelligent, with a high level of familiarity with the workings of the IPCC and knowledge about climate change. Consequently, the figures were designed with an expectation that the audience already has a high level of prior knowledge about the topics being communicated:

*“Its really made for policy makers who are engaged in the process around climate negotiations, climate mitigation, **who would have a good understanding** about global change, climate change issues.” (17)*

*“The mediation that the policy maker might get from these types of people I think will be, at least the ones who I have spoken to, which will probably be half a dozen out of about a hundred dozen that are there, **they had a very high level of knowledge**.” (04)*

*“And so I have always felt, however, this could be a 2-step process. And that is, the big documents and summaries for policy makers, are aimed at an intelligent – an intelligent audience. Governments, so while ministers will have a variety of competencies, they all have technical advisers. So **the cohort in government should be able to understand even quite complex things**. Now this is not a knock-down on ministers. Ministers often have varied portfolios. One day they are the minister for health, the next day they are minister for (inaudible), then they can be foreign office and they can be home office. No-one can be a world expert in every one of those but you would hope the technical advisers to those ministers – and the ministers are intelligent – and with the technical advisers they can interpret what it means basically.” (12)*

*“ ... you know, when you go through all the reviews, I don’t know if you look at the reviews that the policy makers and governments put in, you realise that, **reasonably intelligent people actually**, so we never felt that we were trying to dumb stuff down.” (09)*

*“I wasn’t thinking that we were basically producing them for President Obama, for instance, but come to think of it, sure, there is an educated person who would not shrink from looking at all this. But **very educated people** who actually need to make policy is what I would say.” (05)*

For the scientific community (2.4)

Authors also identified that the report and the figures provide a function for the scientific community, in that they provide a summary or reference source on the science of climate change, enabling individuals to quickly familiarise themselves with areas of research outside of their immediate area of expertise. The scientific community was seen to include both the authors who contribute to the IPCC and scientists who are not part of the IPCC process:

*“So the one function is that it provides a kind of synthesis of the **science for other scientists**. So its, the way I often describe it when I get a question*

about something that is not in my area of expertise but it's a climate science question. My, the first thing that I do is pull the latest IPCC assessment off my shelf and find that topic in there and that usually provides a pretty good introduction to the topic and will point me to the literature if I want to find a few papers that I can read on the topic.” (01)

*“I find it useful to go to those summaries for policy makers to learn about what those two working groups did [referring to working groups 2 and 3]. So I think **for other climate scientists, other scientists, engineers**, it's as much detail as anyone would want to read.” (05)*

*“ ... the SPM is maybe **also for the academic world**. And interesting, erm, the SPM together with the technical summary maybe. If you want to – if you have a student and you say “look, you want to know something about this, well read the SPM, you won't understand anything, read the technical summary you'll understand a bit more, and then you go into the chapter and you pull out the references you need. As such, it's a nice snapshot of our current knowledge.” (14)*

*“Yes, very useful for **other scientists**, very useful I would say for university lecturers.” (15)*

In addition, the inclusion of an author's figure (i.e. a figure they have had a hand in creating and which relates to their research) was assigned as having 'recognition value'. In other words, authors' views and opinions when creating/editing figures and/or reviewing figures may be influenced not only by the remit of the report, but also by a desire to demonstrate their expertise and research work:

*“ ... **authors from individual chapters are very keen to have their research highlighted in the SPM ...**” (04)*

*“There's also the political issues **of giving credit to all those who have done work right**. And that's why you need to show it. But then obviously*

things get complicated and there's too much. It's the same here... all of those dots and hatching and so on..." (18)

As a consequence, while the figures can serve as evidence of 'recognition' among the authors for their contributions to the IPCC reports and therefore be a source of pride, this may come with a risk of including more complexity than may otherwise be needed:

*" ... what I also think that we have to appreciate that **the report is a slightly self-serving enterprise for us as scientists**. I mean you work on different parts of the big climate change issue, and we all think the work we do is the most important of that, so we always try and get our particular thing into... firstly into the report and then in the report try and get it up there, and whether it comes in figures or whether it comes in words, but.. so I think there is **a lot of pressure from the scientists to make all the diagrams too complex**. Because they want to put their own particular thing on them."* (10)

*" ... of course, **everybody would like to bring forward a figure to the SPM, and see their work reflected in the top level document**. But of course there is limited space."* (02)

*"Many of **the people who will judge us for doing these figures are other scientists**, so we want **other scientists to like our figures**. Maybe even more than we want the policy makers to. We don't mind if the policy makers don't like our figures; they should understand what's in there, but this is not about making figures that people will appreciate, - it's making - , it's passing the information."* (08)

*" ... you get people that are so expert that they don't see any more just **how complex their figure becomes**."* (08)

*"If you had them [referring to the figures] in your research paper **you would be very proud of them**. They all maintain a strong consistency.*

They've been checked a million times by world leading experts. You can have very high confidence in them.” (15)

Inaccessible without expert guidance (3)

This theme identifies that non-experts, i.e. individuals who do not have a certain degree of prior knowledge about climate change issues, are expected to find the figures difficult to understand in a stand-alone format, thereby limiting the accessibility of the visuals:

*“The sheer amount of information, the different categories, so you go from the drivers to the quantification, the uncertainty the numbers, the level of confidence, the time information. It’s all in there. And it’s great, but it requires, I mean **a person, who comes to that for the first time just can’t digest it, ...puts it away.**” (02)*

Three sub-themes were identified that demonstrate authors’ beliefs in how the presentation of the figures influences their level of accessibility, which are now summarised. As a consequence of a desire to maintain scientific rigour and to make efficient use of space within the report, many of the figures are thought to be visually complex (**‘visual complexity’**). However, authors are aware that different types of readers of the report have different levels of ability to unpack and interpret this visual complexity (**‘visual formats’**). As a consequence, authors highlight that non-experts need the support from experts to understand and use the information presented in the figures, i.e. to ‘translate’ the information so that it is more understandable (**‘expert explanation needed’**).

Visual complexity (3.1)

A number of the SPM figures were described as being *visually* complex as a consequence of including a high level of informational complexity – i.e. a view that complexity in information largely corresponds with complexity in the representation of that information. Further, visual complexity was also associated

with the need to make efficient use of space in the SPM document. This was articulated in two ways, firstly as a preference for representing information in words in the text of the document, and secondly as a way of combining multiple lines of information within a single figure:

*“Well I mean **obviously they convey more information that you could perhaps have, that you could convey in words**, and not only in terms of the length, but also in terms of the content – I mean you could describe each of these figures in words of course, but it’s actually quite hard for people to follow that if you give people a whole long list of, you know when it was warmer, when it was colder, when it was warmer. I think the reader can grasp, so not only about conveying information, but conveying it more efficiently ...” (04)*

*“Well, there's lots of problems and one of them really is that **if you want to include that robustness, then it makes the figure complicated**. Because you need to show seven different lines.” (18)*

*“And at the end I think we really have the problem of ... **I mean this is highly complicated** multiple different parts, bars, error bars.” (17)*

*“Well, I also think they serve a role in terms of compactness because if you can save a thousand words for every figure you can you can save a lot of text. I say that flippantly, but they do I think **provide a means of conveying a lot of complex information** in a fairly concise and efficient manner.” (01)*

Although visual complexity was expressed as being desirable from the perspective of making good use of available page space, it was also acknowledged as coming with a potential cost of making the information more difficult to understand and interpret:

“On one hand, I admire it because it is such a 'tour de force' of packing information into one 10 by 10 centimetre box, but on the other hand it

takes a lot of work to crack that figure open and really understand it.”
(10)

“The only problem with this figure – and it’s a terrible figure at one level – is you can hardly see the size of – the size of each diagram is very small.” (12)

“So that’s why these figures are complicated because we want to pack two thousand pages of information in ten figures.” (08)

“I understand why it’s there, but it’s... it clutters the figure up, it makes someone looking at it go ‘well which line am I supposed to look at?’ and you know if you read the caption you could, you would eventually realize why the coloured ones are above the solid one is because there’s other greenhouse gases other than CO₂ and that’s their effect.” (01)

*“It’s a lot of different information. There is temperature plots, and also there is ocean heat content plots. There is different curves – you don’t know what they are – you have to read them. Way down the bottom there is – “oh, they’re models. Models – what’s a model?”. “What does that mean?” – a model using natural forces – this is an attribution figure, and – oh then there’s sea ice and a couple of plots that you can’t tell if they’re different from all the other plots. It failed. **I think it’s just way too busy and you have to spend five minutes looking at it and then reading the caption to figure out what it is.**”* (05)

Visual formats (3.2)

Related to the awareness of the presence of visual complexity, was an awareness that certain visual formats - i.e. the type of graph used and visual features to represent certain features of the data – also influence individuals’ abilities to comprehend the presented information. Figures that used sophisticated visual representations were thought to be difficult for non-expert audiences to understand:

“Multi-panel figures are simply too complicated for general public, or for a school, or for even for politicians.” (17)

“If you don’t know anything about statistics, you won’t know why there’s a big white pink cloud around the band, and why there’s a big blue band – you know, what does that mean – what’s the thickness of the band. It’s just too much information.” (05)

“This is a complicated one because it’s two quantities against each other, time running as one of them, as one of the variables. I think this one is easy to look at, but it’s very diff...it might not be that easy to understand what actually is behind the individual lines and the different colours.” (17)

In contrast, simpler visual formats were thought to be easier for people to understand in an intuitive way. In particular, authors believed that graphical formats that are commonly encountered (i.e. familiar) are generally easier for people to comprehend, for example time-series graphs:

*“The most clear, are in general the time series. **The time series people understand better in general.** All time series. They observe then the projections; in general people understand easiest.” (03)*

*“I think that the one I think is probably the best design and communicates most effectively is this sea level figure. **I am attracted to it because it is simple.**” (01)*

*“I mean if you look at S1, **it’s in a way simple and it fits the eye, the eye can easily read it.**” (16)*

*“So... and in fact we found things like **they understood bar charts quite well, they did them at school, so bar charts occasionally work quite well.**” (10)*

*“I think so – yep, time series, quantities, I think that’s pretty straightforward, **you know how to read time series.**” (17)*

Expert explanation needed (3.3)

As a consequence of information complexity, visual complexity and the use of sophisticated visual formats, authors emphasised the importance of explanations to accompany the figures to aid people in understanding their content. The captions of the figures and the text of the report were identified as being key sources of information to enable people to understand the figures, although they also acknowledge that this places an onus on the reader to take the time to read this information:

*“So most of them are capable of understanding this **if they read some text as well.**” (07)*

*“ ... that is a comprehensive way of providing a lot of regional information and information... that one yes... from different components of the climate system. Having said **that it cannot be understood without explanation in the text.** And it’s there. So there is sufficient information to understand it when you read the text ... ” (07)*

*“ ... I think **the figures cannot be looked at in separation of the text and the headlines.** And the sequence in which we present them.” (02)*

*“I understand why it's there, but it's... it clutters the figure up, it makes someone looking at it go ‘well which line am I supposed to look at?’ and you know **if you read the caption** you could, you would eventually realize why the coloured ones are above the solid one is because there's other greenhouse gases other than CO2 and that's their effect”. (01)*

*“Basically I would say none of those figures is understandable to a non-expert without some additional information. It either needs to be **a caption explaining what is shown** because this is relatively technical information or **maybe not even that is enough.**” (18)*

Furthermore, authors emphasised the need for experts to provide verbal explanations to policy makers to assist them in understanding and interpreting the complex information presented in the figures:

*“It doesn't do it in a very clear manner I don't think -- it tries to pack too much stuff in in a way that **requires a lot of explanation**, so I spent a lot of time explaining this figure to people.” (10)*

*“So if you put it out there you let it sit there for a while, **you explain the axes**, you explain what it means, most people get it.” (11)*

*“And I always... this figure you always have to complement with additional information. So with this figure for example, what I always do is kind of **give the example for policy makers**, ‘so now, from such a figure you can read off, what does it mean if you want to stabilise temperature at 2 degree’. ‘What does it mean in terms of emissions?’ So I think it's not the figure that you can put up and then, ‘ok time series goes up’ you don't need to say much. **Here you need to provide a lot of explanations.**” (17)*

*“So even with policy makers – I think this is a key part – is one, when they finally get approve these documents in a plenary, quite often there would be a presentation, especially of the key figures, so when they make a decision, there is at least **some explanation other than the background document.**” (12)*

*“**I can explain most of these pretty simply** – you know two minutes of description and people will get it. A very, very wide audience but it's extremely... as a tool **just on their own without additional explanation they are probably not very useful communication tools.**” (15)*

Discussion

This study identifies that because of maintaining scientific rigour and providing information to primarily expert ‘technical analysts’ (rather than policy makers per se), many of the figures of the AR5 Working Group 1 SPM (IPCC, 2013a)

contain and represent complex information. It further identifies that a secondary audience for the figures is the scientific community. Consequently, IPCC authors are aware that non-experts may find many of the figures difficult to understand. Therefore, authors emphasise the importance of providing explanations for the figures so that non-experts can understand them (Figure 6). These contextual factors in the production and communication of the figures sheds light on how the figures are *intended* to be used.

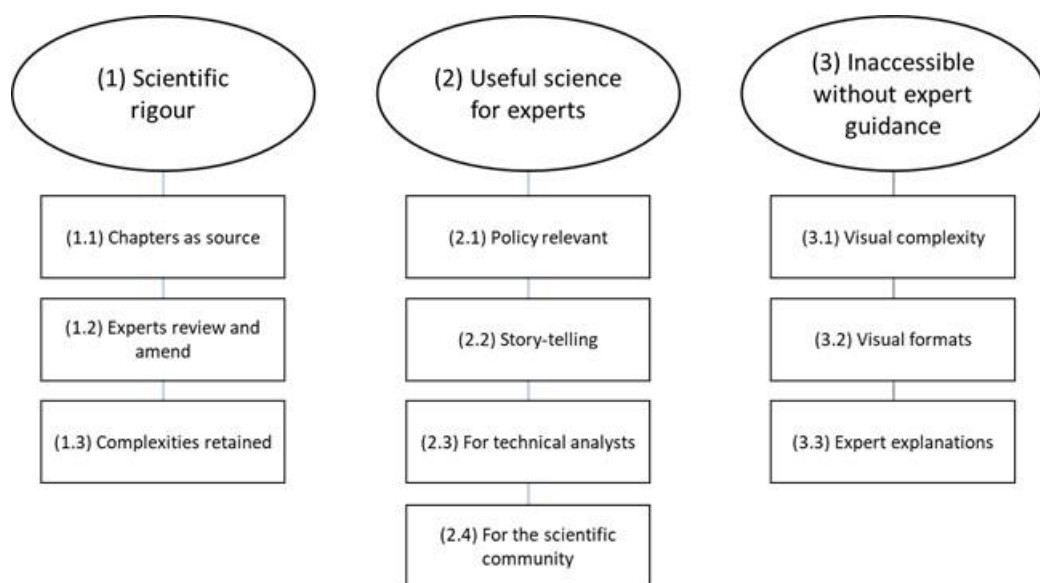


Figure 6. Summary of themes and sub-themes.

The nuance of the types of individuals who are the primary audience of the IPCC SPMs is an important one. The SPM target audience has been taken to include policy-makers in general, elected representatives, media and academics from other subject disciplines (Yohe & Oppenheimer, 2011; McMahon, Stauffacher & Knutti, 2015). These audiences are more varied than just the ‘technical analysts’ that the authors describe. Although critiques about the level of effectiveness of the communication of IPCC reports to such varied audiences may be justified, they largely ignore the technical analyst and intended use of the figures. Conversely critiques that the SPM is a primarily a summary for experts,

rather than broader decision-makers in society (Black, 2015), is consistent with IPCC authors' views expressed in Study 1.

Furthermore, the findings of this study shed light on *why* the authors take this approach of targeting the SPM figures to experts. For figures where maintaining scientific integrity means that they cannot be easily simplified, IPCC authors have the opportunity to *explain* the figures with the technical analysts. IPCC processes have built into them plenary sessions which provides a forum for the authors and analysts to discuss and agree the final content (IPCC, 2013b). Therefore, within this well-defined setting, the IPCC outputs may well achieve their communication goal of communicating policy-relevant information about climate change as the technical analysts will interpret the information in the reports for relevant policy makers.

However, in more broad communication settings beyond the formal IPCC process, the comprehension of the figures by non-experts may be highly dependent on the *availability* of experts to explain the figures to them. Such experts may not always be to hand. Indeed, in the context of the use of the reports by governments, the IPCC authors highlighted that capacities of climate expertise can vary substantially across countries, with some having relatively few climate change science experts. This therefore provides one explanation why on the one-hand IPCC reports are highly regarded by governments – i.e. because from a process perspective government representatives are involved in their production, but on the other hand communication of IPCC outputs, including figures, receive criticism, i.e. because outside of this setting unless additional explanation is provided they can be inaccessible. Here, how the figures are *intended* to be used does not appear to match up with how the figures are *actually* used across a range of contexts.

IPCC authors are aware of these wider contexts (IPCC, 2016), so why is there this mis-match? The current study suggests that authors place a high benchmark for scientific rigour, which from their perspective then limits the extent to which figures can be simplified to be made more understandable. However, is maintaining scientific rigour incompatible with improving the

accessibility of the SPM figures? In other research disciplines, traditional approaches to scientific data visualisation have successfully been enhanced to improve their communication effectiveness while maintaining a high degree of scientific accuracy and detail, for example, in the diagnosis of coronary artery disease (Borkin, et al., 2011) and identifying the threat of storm surges (Sherman-Morris, Antonelli, & Williams, 2015). Such successes have involved collaboration between subject experts and communication experts. However, in the context of the IPCC SPM figures in the present study, authors identified that they themselves were responsible for creating and editing the SPM figures. There was virtually no mention of communication experts, data visualizers, cognitive scientists, psychologists or other specialists being involved in supporting or collaborating with the authors in the construction of the figures. This raises the possibility that the communication of the SPM figures could be enhanced if communications expertise can add new insights into how the climate change data might be better presented to support understanding.

Study 2: Perceived ease of comprehension of IPCC figures

While the interviews (Study 1) identified that the IPCC authors believed that non-experts may experience difficulties in understanding the IPCC AR5 Working Group 1 SPM figures and that they would need accompanying explanations from experts, it may be that the level of comprehension difficulty varies across the SPM figures. Some figures may be more difficult to understand than others. Furthermore, the interviews only provide insights as to the IPCC authors' expectations about non-expert readers, rather the perspectives of non-experts themselves.

On the one hand, IPCC authors may have a good understanding of which figures non-experts find easier to understand and which they find more difficult to understand. Many of the authors have first-hand experience of communicating the work of the IPCC through outreach activities in which they present and explain the reports to various audiences (IPCC, 2016). However, on the other hand, IPCC authors' familiarity with and knowledge about the figures might lead them to erroneous assumptions about which figures non-experts may find easier or more difficult. It has been found that individuals often assume that others have a similar level of knowledge to themselves (Nickerson, 1999). For example, people who have knowledge of the outcome of an event overestimate what their level of knowledge would have been without knowledge of the outcome, and similarly make overestimations about the knowledge held by others who lack knowledge of the outcome (Fischhoff, 1975). Furthermore, greater expertise is associated with worse performance in predicting the time it takes non-experts to complete complex tasks (Hinds, 1999). Such effects are known as the 'curse of knowledge' (Camerer, Loewenstein, & Weber, 1989), which IPCC authors might equally be prone to.

In Study 1, IPCC authors often highlighted particular figures as being very policy-relevant, suggesting that some figures may be more important to communicate to policy makers than others. Consequently, important figures might be designed such that their information is particularly easy to understand – i.e.

such that policy-makers can quickly grasp the information to make use of it. Conversely, it might be that policy-relevant figures represent more complex concepts and ideas, which in turn could be associated with more complex visual representations. For example, observations of indicators of climate change, such as historic temperatures, may be easier to communicate than climate model outputs for future emissions scenarios. However, the later may be more relevant to supporting policy-makers in making decisions about the future.

To investigate experts' beliefs about non-experts, and non-experts' views about the comprehension difficulty of climate science figures, the present study employed a sort task exercise using all ten IPCC AR5 Working Group 1 SPM figures, which was conducted with IPCC authors (experts) and undergraduate students (non-experts). Given that IPCC reports are designed to provide policy relevant information to support decision-making in national governments (IPCC, 2013b), ideally policy-makers working in government, many of whom may be non-experts, would also have been sampled as part of this study. However, policy-makers are a difficult to reach audience to take part in research studies (Burnham, et al., 2008) and therefore university undergraduates provide an alternative, convenient, non-expert sample. It is predicted that experts may hold beliefs about non-experts' comprehension of figures of climate science figures that are different to the actual views held by non-experts, consistent with the 'curse of knowledge'.

Furthermore, the present study also sought to elicit which figures are considered by IPCC Working Group 1 authors (experts) to be most important to inform future climate policy. IPCC assessments are intended to be policy relevant to support decision-making in society (IPCC, 2016). Figures that represent potential futures (i.e. scenarios) might be more important in terms of future climate policy than figures that represent past observation of climate data, because they relate to future choices. However, they may also be more complex. Hence it is predicted that there may be a positive association between the importance of figures to inform future climate policy and their difficulty of comprehension as ranked by non-experts, i.e. more important figures are predicted to be more difficult for non-experts to understand.

Method

To investigate beliefs about the comprehension difficulty of the ten IPCC WG1 SPM figures, a quasi-experiment was conducted with experts and non-experts via a sort-task. First, to see if experts' beliefs of non-experts was aligned with non-experts' actual beliefs, IPCC authors were invited to sort the ten figures based on how they thought university undergraduates would order them in terms of their ease of comprehension. Further, university students were asked to sort the same figures based on their views of their ease of comprehension. The independent variable was therefore the level of expertise, either expert (climate scientists) or non-expert (undergraduate students). The dependent variable was the perceived comprehension ease/difficulty of the figures for non-experts, provided by the ranked order of the figures.

Second, to investigate if experts' beliefs about undergraduate students were similar to their beliefs about policy-makers, climate scientists also sorted the figures based on how they thought policy makers would order them for ease of comprehension. If experts hold similar beliefs for both groups, then undergraduate students might be a reasonable proxy for policy makers when assessing comprehension of climate science figures.

Third, to investigate if there was an association between the importance of the figures and beliefs about their ease of comprehension, climate scientists also sorted the figures based on their importance to inform future climate policy.

Participants

Thirty-eight undergraduate students at the University of East Anglia and eighteen climate change researchers who contributed to past IPCC Assessment Reports took part (all but one of whom were authors to the IPCC Fifth Assessment Report). The sample of climate change researchers were the same individuals who took part in the research interviews (Study 1), of which three were female and 15 were male. University students received course credit or a nominal payment for their participation. Of the university students, 27 were female and 11 male; mean

age was 21 years (range 18-30 years). The majority of the undergraduate students were studying psychology – none were studying environmental sciences.

Materials

The ten figures from the IPCC Fifth Assessment Report Working Group 1 Summary for Policy Makers (SPM) were individually printed in colour on portrait A5 (148mm x 210 mm) card, with each figure having a maximum dimension of 130 mm wide or 190 mm high (Figure 5). Figures were presented with titles, but without captions. On the reverse of each card was a two-digit alphanumeric code to aid data recording.

Procedure: undergraduate students

Undergraduate students took part individually in a quiet room following completion of a separate study (Study 5, Chapter 3). Participants were seated at a desk and were provided with the following instructions to read, *“You will be given a set of 10 cards. Each card will show one or more graphs or diagrams. You will be asked to take a few minutes to look at the contents on the cards – as you do, try to work out what you think the graphs and diagrams are trying to show. Then, please sort the cards in order from the one that you find the easiest to understand (rank 1), through to the one that you find the hardest / most difficult to understand (rank 10).”*

The ten cards were then spread out in a random order in front of the participant and they were then asked to order the cards in a line with the easiest on their left and most difficult on their right. There was no time limit to the task, but participants typically took approximately 3-4 minutes to decide on an order. Participants were then debriefed.

Procedure: climate change scientists

Climate change scientists took part individually, either face-to-face in a quiet room, or remotely via video-conference. The sort tasks were interleaved with interview questions (as reported in Study 1).

Participants were asked to spread out the ten cards in a random order on the desk in front of them. The first task asked them to, “*Rank order the figures from the one you think university undergraduates without climate science training would find easiest to understand through to the one that you think they would find the most difficult to understand.*”. The second task asked them to, “*Rank order the figures from the one you think policy makers would find easiest to understand through to the one that you think they would find the most difficult to understand.*” The final sort task asked them to “*Rank order the figures based on their importance to help inform future climate policy, from the one you think is the most important through to the one that you think is least important.*”

There was no time limit to any of the sort-tasks. Participants typically spent 2-3 minutes completing each task. Interview questions interleaved between each sort task. At the end of each task, cards were collected up, and before the next task the cards were re-shuffled and spread out in a random order. Participants were debriefed after the third task.

Results

Is there a mis-match between experts’ and non-experts’ rankings?

To check whether there was agreement in rankings within each group, concordance among IPCC authors’ rankings and concordance among rankings made by undergraduate students was assessed. There was strong concordance among undergraduates, $W = .473$, $\chi^2(9) = 161.63$, $p < .001$, $n = 38$, and strong concordance among scientists, $W = .566$, $\chi^2(9) = 91.76$, $p < .001$, $n = 18$.

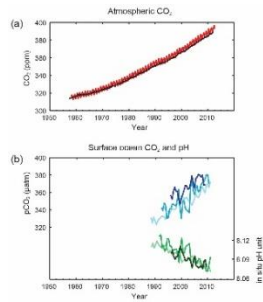
Mean ‘difficulty’ rankings for each figure were then calculated from the undergraduates’ rankings, with low ranks representing figures that were easier to comprehend, and higher ranks representing figures that were more difficult to

comprehend (Figure 7). These mean rankings then provided a criterion ranking against which the strength of association with IPCC authors' rankings about undergraduate students was assessed. There was a weak positive correlation between the ranking of the figures by scientists and the undergraduate criterion ranking, $Tc = .168$, $p = 0.004$ (two-tailed). Hence there was no evidence to suggest a mismatch between IPCC authors' expectations of which figures undergraduate students may find difficult and the undergraduate students' views of these same figures.

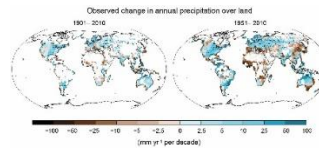
Is there agreement between experts' beliefs about policy-makers and expert's beliefs about undergraduate students?

One expert considered the ten figures to be equally understandable by policy makers and did not provide a rank order for the second sort-task. Among the remaining 17 experts, there was strong concordance among expert's rankings for ease of comprehension for policy-makers, $W = .503$, $\chi^2(9) = 77.03$, $p < .001$, $n=17$.

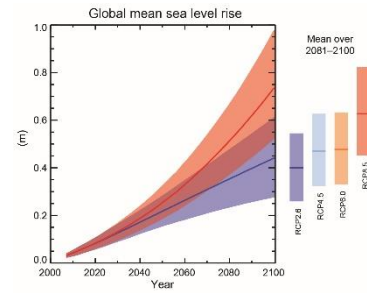
Using each expert's rankings for the ease of comprehension for students to provide a criterion ranking, there was a strong positive correlation between experts' perceptions of undergraduate students and experts' perceptions of policy makers, $Tc = 0.658$, $p < .001$ (two-tailed). Hence, there was a high degree of consistency within each expert in their ranking of the figures for policy-makers and for undergraduate students.



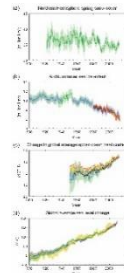
1. Figure SPM.4
 $M = 3.00, SD = 2.25$



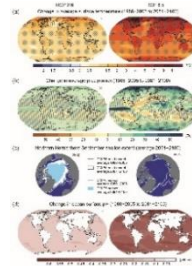
2. Figure SPM.2
 $M = 3.26, SD = 2.02$



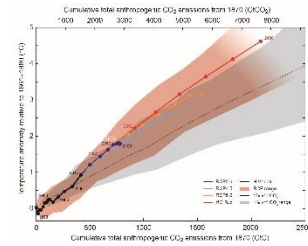
3. Figure SPM.9
 $M = 3.76, SD = 2.36$



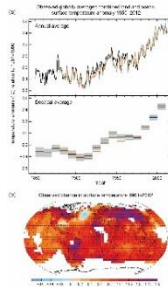
4. Figure SPM.3
 $M = 4.74, SD = 2.22$



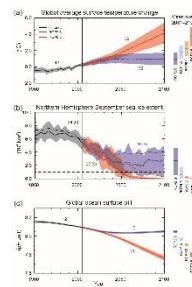
5. Figure SPM.8
 $M = 4.92, SD = 2.14$



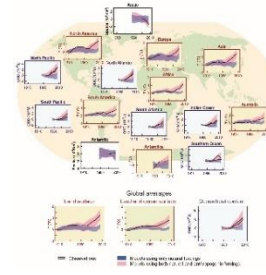
6. Figure SPM.10
 $M = 5.34, SD = 2.29$



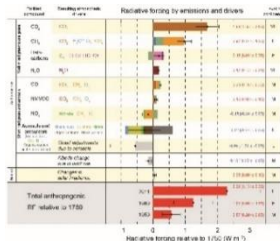
7. Figure SPM.1
 $M = 5.55, SD = 2.42$



8. Figure SPM.7
 $M = 7.05, SD = 1.65$



9. Figure SPM.6
 $M = 7.95, SD = 1.85$



10. Figure SPM.5
 $M = 9.42, SD = 1.43$

Figure 7. Rank order of IPCC AR5 Working Group 1 SPM figures based on their perceived ease of comprehension across all undergraduates – figures are shown

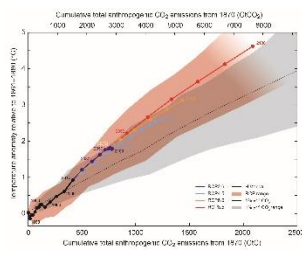
from easiest (rank 1) to most difficult (rank 10). Mean and standard deviation of the ranks are provided under each figure.

Is there an association between the importance of the figures to inform future climate policy and experts' beliefs about their ease of comprehension?

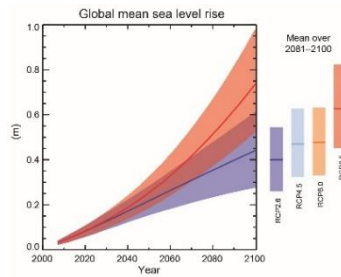
There was strong concordance among expert's rankings of the figures for their importance to inform future climate policy, $W = .517, \chi^2(9) = 83.78, p < .001, n = 18$ (Figure 8).

Using each scientist's rankings for the perceived ease of comprehension of the figures for *policy makers* as a criterion ranking, there was a no significant correlation between scientists' beliefs about policy makers' ease of comprehension and the perceived importance of the figures to inform future climate policy, $Tc = -.064, p < .722$ (two tailed).

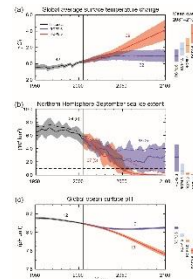
Similarly, using each scientist's rankings of the ease of comprehension of the figures for *undergraduate students* as a criterion ranking, there was a no significant correlation between scientists' beliefs about undergraduates' ease of comprehension and the perceived importance of the figures to inform future climate policy, $Tc = -.067, p < .754$ (two tailed). Hence, across the set of ten figures, there was no evidence to suggest that figures that are perceived as being important (as judged by experts) tend to also be perceived as more difficult for non-experts to understand than less important figures.



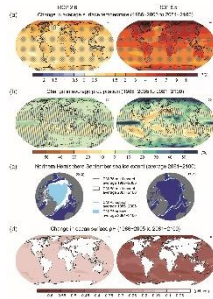
1. Figure SPM.10
 $M = 1.50, SD = 1.17$



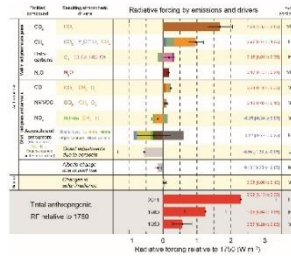
2. Figure SPM.9
 $M = 3.33, SD = 1.05$



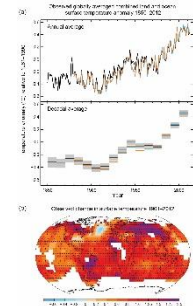
3. Figure SPM.7
 $M = 4.06, SD = 2.04$



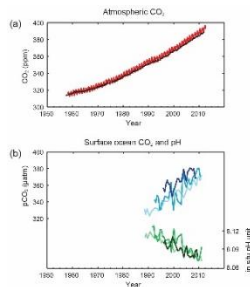
4. Figure SPM.8
 $M = 4.94, SD = 1.96$



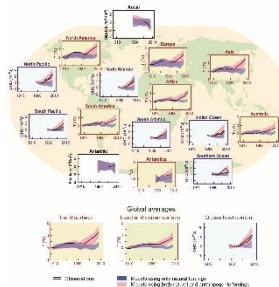
5. Figure SPM.5
 $M = 5.00, SD = 2.58$



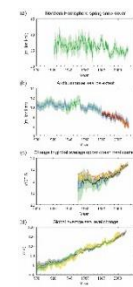
6. Figure SPM.1
 $M = 6.06, SD = 2.37$



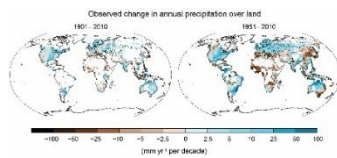
7. Figure SPM.4
 $M = 6.44, SD = 2.39$



8. Figure SPM.6
 $M = 7.11, SD = 2.21$



9. Figure SPM.3
 $M = 8.06, SD = 1.35$



10. Figure SPM.2
 $M = 8.50, SD = 2.17$

Figure 8. Rank order, across all experts, for IPCC AR5 Working Group 1 SPM figures based on their importance to inform future policy. Figures shown from most important (rank 1) to least important (rank 10). Mean and standard deviation of the ranks are provided under each figure.

The figure ranked as being the most important (rank 1) by fifteen of the eighteen expert IPCC authors was Figure SPM.10 (Figure 8), with a mean rank of 1.5. Given the perceived importance of this figure by the experts, the ranks for figure SPM.10 given by experts for the perceived ease of comprehension by undergraduate students, and the ranks given by undergraduate students were then compared in a post-hoc analysis. The mean rank for Figure SPM.10 given by experts about undergraduates was 7.50, whereas the mean rank given by undergraduates was 5.34. There was a significant difference between experts' rankings about undergraduates, and undergraduates' rankings, $U = 164.5$, $p = 0.002$; $N = 46$ (two-tailed); Mann-Whitney mean ranks were 38.36 and 23.83 respectively. Hence, for the most important figure (as judged by experts), undergraduate students ranked it as being easier to understand relative to the other figures, than experts expected them to.

Discussion

The importance of the figures, as evaluated by the IPCC authors, was not associated with expectations about their ease of comprehension. Rather, of the figures that were thought to be relatively important to inform future climate policy, some were expected to be comparatively difficult to understand, while some were thought to be comparatively easy to understand. However, there was a high degree of consistency among the IPCC authors that Figure SPM.10 was the most important figure to inform future climate policy. This was consistent with the views expressed by the authors in their interviews (Study 1), in which this figure was often described as a 'decision-making tool' and therefore very 'policy relevant'. For this figure, there was a mismatch between experts and non-experts, with experts expecting the figure to be comparatively more difficult for non-experts to understand than non-experts perceived it to be.

For the full set of ten IPCC AR5 Working Group 1 SPM figures, there was no evidence to suggest a mismatch between IPCC authors' expectations about which figures university undergraduates might find easy/difficult to understand and undergraduates' actual beliefs. Furthermore, IPCC authors' perceptions were

similar when making judgements about undergraduates and policy-makers. These data suggest that IPCC authors are sensitive to which figures non-experts might have difficulty in understanding. Further, across the set of ten figures the specific type of non-expert (undergraduate student or policy-maker) does not influence experts' opinions of which figures are relatively easy/difficult, and that undergraduates may be a reasonable proxy for policy makers, at least from the perspective of IPCC authors.

These findings contradict that expected by the 'curse of knowledge' in which experts typically over-estimate the ability of non-experts (Nickerson, 1999). In terms of the relative rankings, it appears that experts do have a good sense of which figures non-experts are likely to find easy or difficult. It is important to note that undergraduate students' rankings of the figures were based on their *perceived* comprehension difficulty - their *actual* comprehension was not assessed. Therefore, it is possible that the undergraduates might have perceived certain figures to be comparatively easy to understand, but if tested, their actual comprehension may have been poor. For example, some figures might look deceptively easy to understand, or students might misinterpret seemingly straightforward figures. Further to this point, IPCC authors acknowledged that the figures are difficult for non-experts to understand (Study 1) and evidence suggests actual comprehension may indeed be poor (McMahon, Stauffacher, Knutti, 2015).

Indeed, this might explain why figure SPM.10 was ranked comparatively easier to comprehend by undergraduates than experts considered it to be for undergraduates. Authors who are familiar with the figure and the concepts that are represented by it, may judge the relative difficulty of the figure on the basis of how easy/difficult the concepts are to grasp. Conversely, undergraduates lack expert knowledge and may have made superficial judgements regarding the difficulty of the figures based on their intuitions about the visual representations, rather than a meaningful effort to comprehend the information in each figure. Indeed, current theories regarding cognitive processing of data visualisations emphasize the role of prior knowledge alongside bottom-up perceptual processing (Freedman & Shah, 2002; Pinker, 1990). Given that non-experts lack the domain knowledge held by experts, they may therefore rely on surface perceptual

properties of the figures when interpreting them (Shah, 2002). Hence, rankings for ease of comprehension made by non-experts in the present study might have been made by intuitive perceptual judgements about the visual complexity of the figures. This possibility is now explored in Study 3.

Study 3: Visual complexity of IPCC Figures

Study 2 identified a high level of agreement across non-experts in the rank order of the figures for perceived comprehension ease/difficulty, suggesting that these rankings may be based on common criteria. If so, what might these criteria consist of?

By definition, non-experts lack the domain knowledge of experts. In the case of the figure judged to be the most important by experts – Figure SPM.10 – experts believed non-experts would judge it to be relatively more difficult to comprehend than the non-experts actually did. If the undergraduate students did not use expert domain knowledge to make their rankings (in Study 2), they may have relied on perceptual features of the graphs. Indeed, as identified in the interviews (Study 1), experts held views that the visual complexity, often expressed in terms of the information density of a visual, was one reason why some figures are more difficult for non-experts to understand than others. Hence, figures that have a high level of visual complexity may be perceived by non-experts as being difficult to understand. Study 3 therefore asks to what extent is the visual complexity of a data visualisation associated with the perceived comprehension difficulty?

What is visual complexity and how can it be measured?

The complexity of a data visualisation might relate to the complexity of the visual information (i.e. the perceptual features of the data visualisation) or complexity of the referents to which the visual information refers (i.e. the complexity of the concepts that are being represented in the visual). Here the focus is on the *visual* complexity of the information, to explore the *perceived* comprehension difficulty as rated in Study 2.

A number of definitions of visual complexity have been suggested, including the degree of detail within the visual (Snodgrass & Vanderwart 1980), the degree of variation of parts (Heylighen, 1997), and, specifically in relation to

graphs, the number of points plotted, the configuration of these points into perceptual groups, and the consistency of the patterns created by these groups (Meyer, Shinar, & Leiser, 1997). The emphasis on perceptual organisation within these definitions is consistent with the well-established Gestalt laws (Bruce, Green & Georgeson, 2003), which enable visual information to be grouped into meaningful ‘units’ or ‘chunks’. Furthermore, these definitions also emphasise that complexity relates to the amount of perceived variation of the ‘units’ or ‘chunks’. Such perceptual mechanisms are thought to support cognition of data visuals as they enable cognition to be ‘offloaded’ onto perception (Hegarty, 2011).

The conceptualisation of visual complexity has been explored by asking individuals to make subjective judgements about sets of visuals (Moacdieh & Sarter, 2015). Subjective judgements can be elicited by asking participants to rate visuals for complexity against a given criterion, such as “the amount of detail or intricacy” (Snodgrass & Vanderwart, 1980, pp 183), or to classify visuals based on the perceived degree of complexity and to provide verbal descriptions (Heaps & Handel, 1999). Evidence from subjective classification of photographs of real-world scenes suggests that individuals’ perceptions of complexity are based on the quantity of perceived objects, followed by the relationship between the number of perceived objects and their spatial arrangement (Oliva, Mack, Shrestha, & Peeper, 2004).

The relationship between the number of objects and their spatial layout is typically known as ‘visual clutter’ (Peng, Ward, & Rundensteiner, 2004; Oliva, Mack, Shrestha, & Peeper, 2004; Doyon-Poulin, Robert & Ouellette, 2012), which has been defined as an excessive amount of information, or spatial disorganization of information, which impairs task performance (Rosenholtz, Li, & Nakano, 2007). For example, greater visual clutter can result in task errors (Baldassi, Megna, & Burr, 2006) and slow the identification of targets in a visual display (Neider & Zelinsky, 2011). This conceptualisation of complexity as ‘visual clutter’ also draws on the Gestalt principles of spatial organisation (Bruce, Green & Georgeson, 2003), for example multiple data points organised such that they fall on a line may be perceived as a single perceptual unit based on their spatial proximity. Visual clutter may therefore be a useful criterion for

determining the visual complexity of data visualisations, especially for relational displays, in which the spatial arrangement of visual elements is used to demonstrate relationships in the data.

Objective measures of visual clutter have typically be obtained through measures of task accuracy (e.g. Wickens, Nunes, Alexander, & Steelman, 2005; Alexander, et al., 2012) or through reaction times on visual search tasks (e.g. Neider & Zelinsky, 2011; Yeh, & Wickens, 2001). While such measures are useful in evaluating visual clutter in contexts where there are well-defined user-tasks, such as finding a specific visual element in a data visualisation, they may not easily apply to less well-defined tasks. For example, data visualisations are often intended to demonstrate how data provides evidence for a particular finding or conclusion, i.e. for the less well-defined task of ‘communicating a message’. Extracting a message from a data visualisation may involve numerous cognitive processes, such as visual search (Hegarty, 2011) and spatial inferences (Trickett & Trafton, 2006). Furthermore, in less well-defined tasks, task accuracy can be difficult to objectively define.

Computational measures of visual clutter

An alternative way of quantifying visual clutter is through computational models (Rosenholtz, Li & Nakano, 2007). Such models extract statistical properties from the visual image to determine the extent of clutter present and can be validated against subjective assessments of clutter, or objectives measures such as reaction times to visual search tasks. Computational models of clutter are inspired by knowledge of human cognition, for example, modelling characteristics of visual attention (Da Silva, Courboulay, & Estrailier, 2011; Rosenholtz, Li & Nakano, 2007) or perceptual organisation (Rosenholtz, Li, & Nakano, 2007).

A key advantage of computational models of clutter is that they can be easily applied to naturalistic images, and so support scaling of cognitive theories of visual displays to real-world stimuli. Existing models of graph comprehension acknowledge the role of bottom-up visual processes on influencing comprehension (Pinker, 1990; Freedman & Shah, 2002), but are under-specified.

For example, ‘set size’ is a characteristic of the visual display representing the number of objects present and is known to influence performance on visual search (Wolfe & Horowitz, 2004). While counting the number of points plotted in a simple x-y scatterplot can be easily determined, set size does not say anything about the spatial organisation of the data points. Furthermore, in more complex graphs, such as the IPCC figures, it is not obvious what counts as an ‘item’ when calculating set size. Therefore, computational models offer a potential solution to quantify the extent of visual clutter in a complex data visualisation, and so provide a proxy for visual complexity.

Subband entropy as a measure of visual clutter

Numerous computational image-processing based models have been proposed to measure visual clutter in images (for a review see Moacdieh & Sarter, 2015). In the context of measuring the visual clutter of data visualisations, which typically encode meaning via the spatial arrangement of visual features, a computational measure that captures the extent of spatial organisation may offer a potentially useful indicator of clutter, and therefore complexity.

Subband entropy is related to the degree of organisation of a visual scene, which in turn can be conceptualised as the extent to which one part of the visual is predictable from another part (Rosenholtz, Li, & Nakano, 2007). Hence, as visual information becomes more organised, it becomes less cluttered and therefore less (visually) complex. Subband entropy is based on the same principles of JPEG image compression algorithms, in which the image is divided into subbands, each representing different spatial frequencies and orientations, and the amount of entropy within each subband is computed. The sum of the entropy across subbands provides a clutter score, with lower scores representing less visual clutter (Rosenholtz, Li, & Nakano, 2007).

The handling of the spatial characteristics of the visual information by the subband entropy measure may make it a more appropriate measure of visual clutter for relational data visualisations than feature congestion models of clutter (Rosenholtz, Li, & Nakano, 2007; Rosenholtz, Li, Mansfield, & Jin, 2005), which

operate by determining the number of items in a display and the extent to which these make it more difficult to add further items such that they are visually salient. The subband entropy measure of visual clutter has been validated against performance on visual search tasks (Rosenholtz, Li, & Nakano, 2007) and has been used to inform display design for geographic information systems (Wiehr, Setlur, & Joshi, 2013) and human-computer interfaces (Miniukovich & De Angeli, 2015). However, it is not yet known if subband entropy performs well in relation to people's subjective assessments of the ease of comprehension of data visualisations.

The present study / design

To explore if visual clutter is associated with perceived comprehension difficulties, undergraduate students' judgements (rankings) of the comprehension difficulty of ten climate science figures (data as per Study 2) was compared with the degree of visual clutter in the figures, measured by subband entropy.

Method

Undergraduate students' rankings of the ten IPCC Working Group 1 SPM figures collected in Study 2 were re-used in the present study. Hence participants, stimuli and the procedure were as per previously reported.

Computing visual complexity

The ten figures were cropped to the edge of the images (so that clutter was measured on the visual content). Images were exported as jpeg files, with maximum pixel dimensions of 740 pixels high by 540 pixels wide. An index of visual clutter for each figure was then calculated using the subband entropy measure (Rosenholtz, Li, & Nakano, 2007).

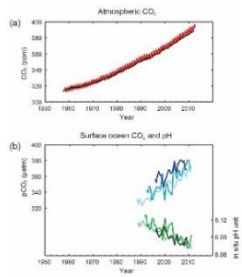
Results

Main analysis was across all ten figures, but due to the heterogeneity of graph format, a sub-analysis was performed for the seven abstract relational graphs, i.e. excluding Figures SPM.1, SPM.2 and SPM.8, which contain thematic maps.

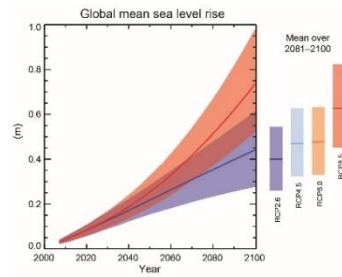
Unlike abstract relational graphs, the spatial organisation and content of thematic maps may closely reflect conceptual organisation because they directly represent visual entities in the world (Hegarty, 2011).

First, the level of agreement between participants in their rankings was checked. There was strong agreement in rankings across all figures, Kendall's $W = .473$, $\chi^2(9) = 161.63$, $p < .001$, and across the sub-set of the abstract relational graphs, $W = .560$, $\chi^2(6) = 127.67$, $p < .001$; indicating participants used similar criteria to order the figures. Mean ranks for perceived comprehension difficulty are as per those reported in Study 2 (Figure 7).

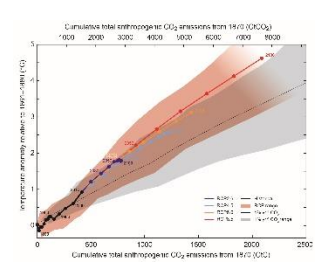
Visual clutter scores, as measured by subband entropy, ranged from 2.743 to 3.981 (higher scores indicate more visual clutter). The figures were ranked using their subband entropy scores to provide a visual clutter criterion ranking (Figure 9).



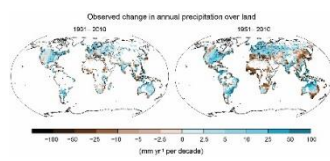
1. Figure SPM.4
Subband entropy: 2.743



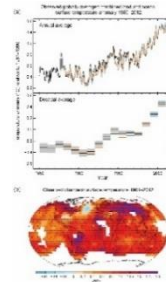
2. Figure SPM.9
Subband entropy: 2.774



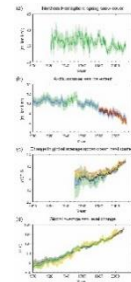
3. Figure SPM.10
Subband entropy: 2.985



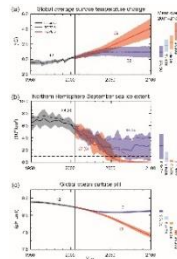
4. Figure SPM.2
Subband entropy: 3.309



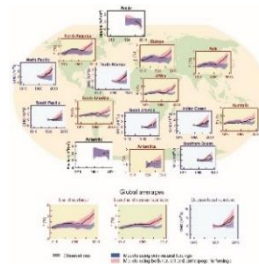
5. Figure SPM.1
Subband entropy: 3.364



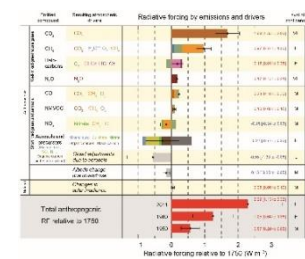
6. Figure SPM.3
Subband entropy: 3.394



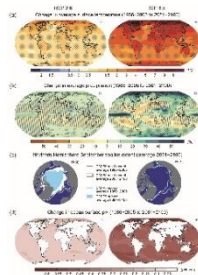
7. Figure SPM.7
Subband entropy: 3.418



8. Figure SPM.6
Subband entropy: 3.533



9. Figure SPM.5
Subband entropy: 3.769



10. Figure SPM.8
Subband entropy: 3.981

Figure 9. Subband entropy scores for each figures, numbered from the least cluttered (lowest subband entropy score) through to the most cluttered figure (highest subband entropy score).

The strength of relationship between visual clutter and perceived comprehension difficulty was then assessed using T_c , (the average of Kendall rank-order correlation coefficients between the criterion ranking and each participant's ranking). Across all ten figures there was a medium to large positive correlation, $T_c = .399$, $p < .001$, and a large positive correlation for the sub-set of the seven abstract relational graphs, $T_c = .622$, $p < .001$. Greater visual clutter was associated with greater perceived comprehension difficulty, with a stronger association between these variables when the analysis was restricted to abstract relational graphs.

Discussion

The study found that perceived comprehension difficulty of the IPCC Working Group 1 SPM figures by non-experts (i.e. undergraduate students) was associated with the extent of visual clutter in the figures, as determined by the subband entropy – a computational image-based model of visual clutter. Figures that contained greater visual clutter were perceived as being more difficult to understand than those that contained less clutter. These results confirm IPCC authors' beliefs that figures that are visually complex will be more difficult for non-experts to understand (Study 1) and provide strong evidence that visual clutter closely maps with the criterion used by non-experts to make judgements about the difficulty of the figures (Study 2). Further, the results further support the potential utility of subband entropy as measure of visual complexity with complex data visualisations (Rosenholtz, Li, & Nakano, 2007).

The stronger association between visual clutter and comprehension difficulty for the abstract relational graphs than the complete set of figures raises the possibility of distinct underlying relationships between clutter and comprehension for different visualisation formats. While the outlines of the continents in a thematic map might create visual clutter as determined by subband entropy, their spatial organisation reflects conceptual organization (i.e. prior knowledge of spatial organisation of continents), making cognition comparatively easy (Tversky, 1997). Indeed, models of clutter that are purely image-processing

based cannot not account for the influence of top-down knowledge held by the reader (Moacdieh & Sarter, 2014). Hence if visual elements create visual clutter, but convey well known meanings, then the clutter may not have a detrimental effect on higher level cognition. Consequently, it may not be appropriate to use purely image-processing based methods to assess visual clutter when making assessments across different types of scientific figures.

In consideration to the abstract relational graphs, given that participants were all expected to have minimal domain knowledge and similar levels of graph knowledge (all were university undergraduates who were not studying climate science), top-down knowledge may have been relatively homogenous. Consequently, concordance across participants in the relative comprehension difficulty of the figures would be expected to lie in differences in bottom-up processes acting on the visual features of the figures. In such contexts, computational measures of visual clutter, such as subband entropy, could be a useful diagnostic tool to assess data visualisations for complexity.

However, across more heterogenous populations, variation in top-down domain and/or graph knowledge (Hegarty, 2011) might be equally or more important in informing comprehension difficulty, where there might be less concordance. Similarly, in real-world settings, user-based factors may also be important to explain differences in perceived comprehension across figures, such as motivation and workload (Moacdieh & Sarter, 2014). For example, it might be that motivated individuals find that certain figures can initially ‘appear’ to be easy to comprehend (e.g. if they have minimal visual clutter), but whose comprehension on further inspection is challenging. Less motivated individuals might simply follow their initial superficial perceptions.

Further research is needed on how visual clutter and perceptions of difficulty might translate to actual difficulty. For example, perceptions of difficulty might influence the degree to which an individual engages with a data visualisation and might be associated with actual difficulty, mediated via reduced self-efficacy (Mangos & Steele-Johnson, 2001). However, if given well-defined tasks and relevant knowledge, readers may then perform well, especially when

task relevant information is made perceptually salient (Hegarty, Canham, & Fabrikant, 2010). For example, comprehension may only be impaired by visual clutter if task-relevant information is not visually salient (Wilkening & Fabrikant, 2011).

In summary, if non-experts are simply tasked to quickly understand a data visualisation, the degree of visual clutter might be a useful indicator of perceptual complexity and perceived comprehension difficulty.

General discussion

The three studies presented in this chapter highlight that communicating societally relevant scientific evidence in data visualisations can be particularly challenging when maintaining a high level of scientific rigour. In the interviews with IPCC authors in Study 1, the need for scientific rigour was seen to result in the creation of figures that contain a high level of informational complexity and which need to be accompanied by expert explanations to enable non-experts to understand them. Furthermore, experts were sensitive to which figures non-experts perceive to be comparatively more difficult than others (Study 2). A high level of visual complexity in the figures, which may be a consequence of retaining a high level of informational complexity (Study 1), was positively associated with non-experts' judgements of their perceived comprehension difficult (Study 3). Together, the evidence indicates that visual representations used by the IPCC are not optimally designed to enable non-experts to understand them in the absence of expert support or guidance.

These findings are in concordance with research findings that non-experts can misinterpret the meaning of IPCC figures (McMahon, Stauffacher, & Knutti, 2015) and that users of IPCC reports consider IPCC figures to be 'low quality communication tools' (IPCC, 2016, pp. 104) where their content is too complex for the needs of policy makers and non-experts. This may be particularly problematic for important figures to inform future climate policy, such as Figure SPM.10 from the IPCC Fifth Assessment Working Group 1 report, which is seen by authors as extremely policy-relevant yet also comparatively difficult for non-experts (including policy makers) to understand (Study 2).

It is well established in the field of psychology that representations that are informationally equivalent are not computationally equivalent (Larkin & Simon, 1987). For example, plotting the same data in line graphs and in bar graphs results in different inferences (Shah & Freedman, 2011). Hence, it is possible that carefully constructed alternative representations of IPCC figures, whose representations align with the key inferences that readers are expected to make from the data, might enhance non-expert comprehension. Furthermore, in

the absence of relevant domain and/or graph knowledge, non-experts appear to be particularly sensitive to bottom-up perceptual features when trying to comprehend complex figures (Study 3). However, when non-experts acquire relevant top-down knowledge, performance when interpreting information with visual displays has been shown to improve (Hegarty, Canham & Fabrikant, 2010; Shah, Freedman & Vekiri, 2005). Therefore, non-experts' comprehension of figures might also be enhanced by considering ways in which prior knowledge needed to interpret figures can efficiently be imparted.

It is important to note that the figures presented to participants in Study 2 contained only the figure and the figure's title, and did not include any accompanying figure captions or other text. Figure captions, and text that refers to figures in the body of a report, may be a source of key information, providing top-down knowledge, to facilitate comprehension of a figure. Indeed, when referring to the need for 'accompanying explanations', experts sometimes referred to the text of the report in this role (Study 2) and captions can support comprehension of figures (Slough, McTigue, Suyeon, & Jennings, 2010). However, the text of IPCC reports score poorly on readability metrics (Barkemeyer, et al., 2016) and captions have been criticised for being lengthy and difficult to understand as a consequence of the figures trying to pack in too much information (IPCC, 2016). Furthermore, figure captions typically do not convey the communication 'message' of the figures, but rather descriptive information to describe referents (Elzer, et al., 2005). Therefore, it is possible that inclusion of the captions or supporting text in Study 2 may not have made much, if any, difference in the perceived comprehension difficulty rankings of the figures by non-experts. In addition, the utility of such information in applied contexts may be limited due to figures and associated text not being presented in close spatial proximity – a factor known to require divided visual attention, resulting in impaired comprehension (Holsanova, Holmberg & Holmqvist, 2009).

Requiring readers to read and understand lengthy and complicated linguistic information to understand data visualisations places the burden of comprehension on the reader (i.e. by top-down knowledge). Conversely, designing data visualisations such that key messages can be easily grasped by the

intended audience(s) places responsibility of effective communication on the creators of visuals. The studies presented in this chapter identify that there is an unmet need to improve comprehension of real-world societally relevant scientific data visualisations. The application of cognitive insights (Chapter 1) to the design and communication of such figures provides a mechanism to achieve this. Furthermore, in doing so, there is the opportunity to advance understanding of cognitive processes involved in graph comprehension in ecologically valid contexts. These aspects are next explored in Chapter 3.

Chapter 3: Supporting spatial inferences

This chapter considers how people make spatial inferences about data presented in time-series graphs. Time-series graphs can convey information about long-term trends and short-term variability and are the most common graph format used in the IPCC AR5 Working Group 1 SPM. In the context of communicating climate change, the primary purpose is to convey information regarding long-term trends of indicators of a changing climate. However, a high degree of short-term variability in such graphs may mask the trend.

Study 4 presents a pilot study exploring what information non-experts spontaneously describe when asked to interpret one of the IPCC AR5 Working Group 1 SPM time-series figures. Verbal interpretations indicate that non-experts often describe the short-term variability in the data and do not always describe long-term trends.

Study 5 then investigates whether comprehension of long-term trends can be supported via a linguistic warning. The results indicate that a warning instructing individuals to identify trends and ignore extreme data points directs visual attention to trend-relevant information (measured using eye-tracking), which then supports improved spatial representations for trends in memory.

Study 6 attempts to replicate the effect of the linguistic warning found in Study 5, and further extends the research by exploring whether linguistic warnings are goal-dependent or goal-independent. The goal-dependent hypothesis proposes that spatial representations of trends will only be supported by a warning that matches the goal, e.g. a warning to ignore extreme data and identify trends. Conversely, the goal-independent hypothesis proposes that *any* warning to identify something and ignore something else enhances attention more generally, which then supports improved spatial representations for trends. Results across Studies 5 and 6 indicate a reliable effect of a linguistic warning in supporting spatial representations for trends. Study 6 also provides some evidence for the goal-independent hypothesis.

Findings highlight that linguistic warnings can increase attentional vigilance and so support people to form robust spatial representations for information that is not explicitly represented in a visual display. Further, the studies demonstrate that relatively simple cognitively inspired interventions, which provide top-down knowledge, have the potential to enhance non-expert's encoding of spatial information, which may then support improved comprehension.

Interpreting trends in data

As identified in Chapter 1, there is the potential to enhance the communication of scientific figures by understanding the cognitive processes involved in their comprehension. In interviews with IPCC authors (Study 1) figures from the Working Group 1 SPM were perceived to be difficult for non-experts to understand. Therefore, generating a deeper understanding of cognition for complex real-world data visuals, such as those used the IPCC, offers an opportunity to gain insights on how to enhance their comprehension.

Although a range of graph formats are used across the IPCC reports, within the IPCC Working Group 1 SPM (IPCC, 2013a), seven of the ten figures contain line graphs, typically plotting time on the x axis and a measured or modelled variable on the y axis. Such figures are used to demonstrate how variables change over time, for example Figure 10.

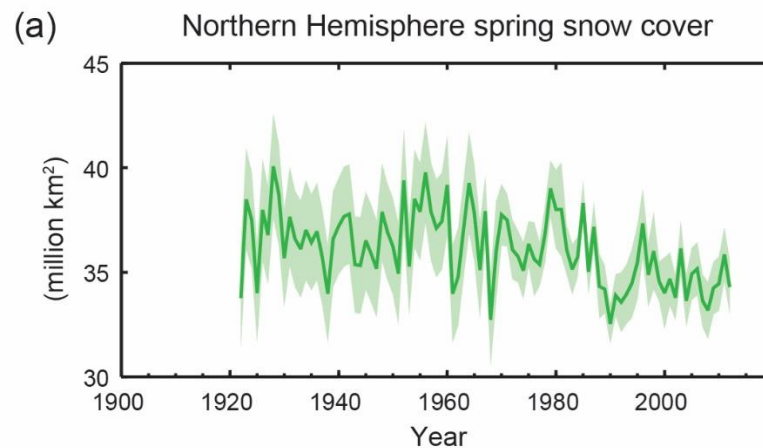


Figure 10. Figure SPM.3, panel a, showing extent of Northern Hemisphere March-April (spring) average snow cover; time-series show annual values, and where assessed, uncertainties are indicated by coloured shading. Reproduced from IPCC, 2013a.

Indeed, line graphs are particularly common across print publications, for example they have been shown to make up more than 50% of graphs in journals, magazines and newspapers (Zacks, Levy, Tverksy, & Schiano, 2002). Furthermore, most line graphs in such publications plot time-series data (Borkin, et al., 2013).

Time series graphs represent data by a connected line through each adjacent data points, resulting in a complex line graph. Broadly, two general spatial characteristics of time-series data can be extracted from a line graph. First, short-term *variability* describes short-term fluctuations in the data, which is explicitly represented in the display (the vertical spread). Second, *trends* describe the general slope of the data across time, typically across the whole data set or significant parts thereof, which must be inferred. Inferring the long-term trends in historical data allows us to interpret underlying relationships. For example, indicators of a changing climate, such as spring snow cover, show underlying *trends* over multiple decades (Figure 10), in this case showing spring snow cover has decreased. There is also short-term variability – the data does not decrease

every year, sometimes it increases from one year to the next. Substantial variability might mask long-term trends. Therefore, in such visualizations, can non-experts efficiently and accurately identify long-term trends? If not, can language support trend identification?

Perception of a complex line

The overall shape of a complex object, such as a complex line graph, is thought to be poorly defined during early visual processing (Wolfe & Bennett, 1997) and may instead be decomposed into parts or ‘chunks’ based on boundaries created by local curvature extrema, defined as points of negative minima in the shape (Hoffman & Richards, 1984). Time-series data that show significant variability have numerous curvature extrema (i.e. trend reversals) creating numerous visual chunks (Figure 11).

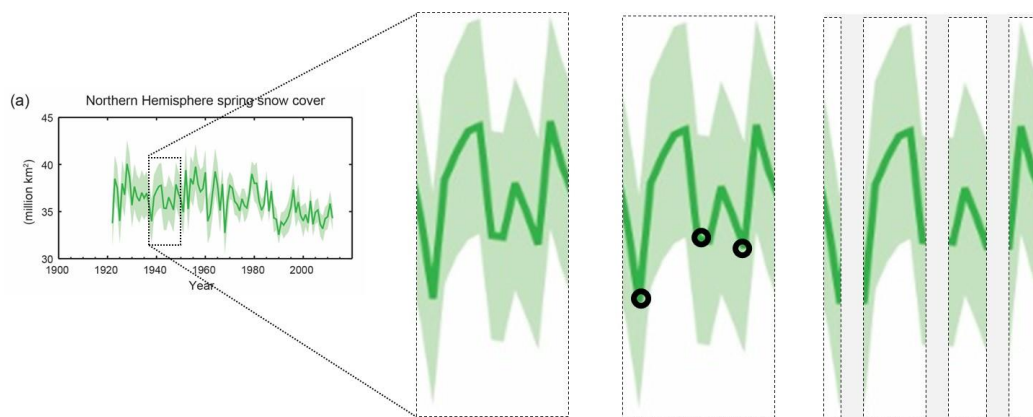


Figure 11. Schematic of how part of a complex line (solid green line) might be decomposed into chunks by segmenting the line at points of local curvature extrema. Black circles in middle box indicate points at which the connected line may be segmented. Right-hand box shows the resulting segmentation. In this example, for simplicity, it is assumed that the area under the line is foreground and the area above the line background, to determine locations of local curvature extrema. Figure shown is Figure SPM.3 reproduced from IPCC, 2013a.

Decomposition of a shape into parts is thought to happen pre-attentively (Baylis & Driver, 1995, Driver & Baylis, 1995), and the salience of any given part being dependent on the relative size of the part to the complete object, the extent of the protrusion and the turning angle of the concavity (Hoffman & Singh, 1997). Hence time-series graphs showing variability may automatically be chunked into component parts, some of which may be more salient than others. Short-term variability is therefore explicitly represented in the graph, and is directly perceivable.

Conversely, identifying the long-term trend of the data may require integration of chunks using spatial processing (Freedman & Shah, 2002; Carswell, Emery, & Lonon, 1993). Spatial processing in this context refers to holding spatial information in working memory and/or performing spatial transformations on mental representations of objects, both of which are thought to be common when conducting tasks with complex data visuals (Trickett & Traflet, 2006).

Evidence suggests the need for spatial processing when interpreting complex line graphs. The number of trend reversals in line graphs has been associated with increased study time, and with increases in local content in verbal and written interpretations at the expense of global content (Carswell, Emery, & Lonon, 1993). Furthermore, line graphs with a large degree of short-term variability are associated with poorer performance on an aggregate judgment task (Correll, Albers, Franconeri, & Gleicher, 2012). Therefore, although simple line graphs may be efficient representations to convey meaning of trends (Zacks & Tversky, 1999), it is not known whether this is also true for more complex line graphs that contain short-term variability.

Study 4 (pilot study): Do non-experts describe trends in an IPCC figure?

The aim of Study 4 (pilot study) was to characterize difficulties, if any, in trend interpretation by asking participants to look at and then describe a real-world time-series graph that contained an underlying long-term trend as well as substantial short-term variability. Asking people to describe what they think a graph shows can identify which information is salient and encoded (Hegarty, 2011; Shah & Carpenter, 1995).

To see if people correctly identify long-term trends from time-series graphs that also show significant short-term variability, verbal descriptions were collected from individuals exposed to a real-world graph showing such characteristics. The graph chosen (Figure 10) shows data for Northern Hemisphere spring snow cover extent between 1922-2012, published by the Intergovernmental Panel on Climate Change (IPCC, 2013a) which is one of the figures from the AR5 Working Group 1 SPM. The data indicate a significant downward trend over the whole time-period, together with substantial inter-annual variability. In the text of the SPM, the authors indicate that snow cover extent has decreased since the mid-20th century (IPCC, 2013a), suggesting that this message is an important communication goal. Given that the short-term variability is explicitly represented in a complex line graph, whereas the long-term trend is not, it is predicted that the majority of individuals will describe short-term variability, but may not describe long-term trends.

Method

Participants

Twelve undergraduate students (10 female, two male) from the University of East Anglia took part in the pilot study in return for course credit or a nominal

payment. Their average age was 21 years (range 19–29 years). None of the participants were studying environmental sciences.

Apparatus and Materials

The target stimulus consisted of Figure SPM.3a from the IPCC SPM (IPCC, 2013) (Figure 10). The stimulus were presented on a TFT LCD monitor (51cm x 29cm), set to 1280 x 720 pixels. Eprime Version 2.0 (Psychology Software Tools Inc., Sharpsburg, USA) was used to control stimulus presentation and record data. Verbal responses were captured via a headset microphone.

Procedure

Participants were instructed that on each trial, they would be shown a graph or diagram to study and would then be prompted to “describe what you think the graph is trying to show”. The trial in the present study was presented to participants embedded within another study (Study 5). Participants were therefore shown a visual prompt indicating that they should study and prepare to describe the next graph they see. The graph was presented for 15 seconds, during which participants simply looked at the figure. Participants then saw a ‘Now describe’ prompt and the same figure re-appeared on the screen. The figure remained on screen until the participant completed their verbal response (indicated by pressing the spacebar on the keyboard) or until a maximum time limit of 45 seconds was reached (Figure 12).

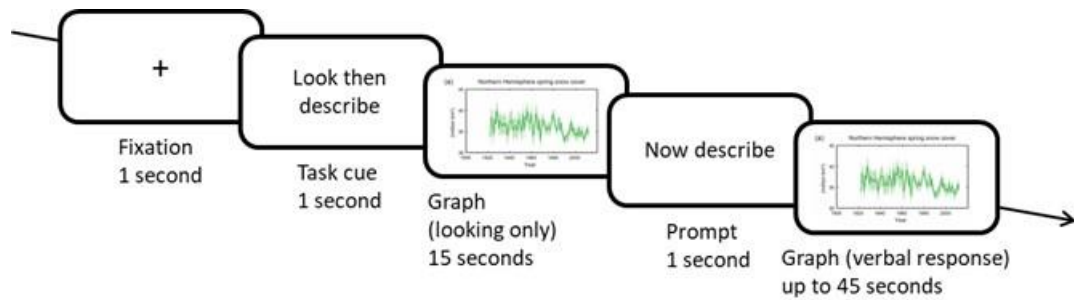


Figure 12: Presentation of experimental trial.

Coding

Verbal descriptions were coded to assess the presence (coded as '1') or absence ('0') of the following characteristics: (a) the data represent changes in snow cover over time; (b) a general downward trend; (c) a downward trend between ~1960 and ~2012; (d) short-term variability/fluctuation. Coding criteria are shown in Table 3. Inter-rater reliability across all aspects and all coding was $K = 1.000$, $p < .001$ (i.e. complete agreement).

Table 3. Study 4 coding criteria for the four characteristics.

Characteristic	Characteristic phrases used to code
(a) the data represent changes in snow cover over time	Refers to ‘snow cover’ and refers to time in the same utterance, such as ‘over time’, ‘over years’, ‘between ~1900 and ~2012’.
(b) a general downward trend	Refers to the plotted data, as a whole, showing a downward trajectory, such as ‘going down’, ‘decreasing’, ‘decline’ and does not tie this description to a specific time period, or explicitly refers to the whole time period.
(c) a downward trend between ~1960 and ~2012	Refers to the plotted data as showing a downward motion, such as ‘going down’, ‘decreasing’, ‘decline’ and ties this description to a specific time-period congruent with the period of the graph representing ~1960 and ~2012.
(d) short-term variability/fluctuation	Refers to the plotted data as showing variability, such as ‘peaks and troughs’, ‘goes up and down a lot’.

Results

All twelve participants correctly identified that the data represented changes in snow cover over time (Table 4). Five participants (42%) described some form of downward trend (either a general trend and/or a trend ~1960 and ~2012), whereas seven participants did not describe any form of a trend (58%). Taking those who did describe a trend and those who did not describe a trend as two groups, the likelihood of describing the short-term variability was then compared. Of those who described a trend, only one (20%) also described short-term variability, compared with five of the seven participants who did not describe a trend (71%) ($p = .01$, Fisher’s Exact Test).

Table 4. Study 4 frequency of the number of individuals who verbally described each characteristic.

Characteristic	Frequency count (percentage)
(a) the data represent changes in snow cover over time	12 (100%)
(b) a general downward trend	5 (42%)
(c) a downward trend between ~1960 and ~2012	1 (8%)
(d) short-term variability/fluctuation	6 (50%)

Discussion

These pilot data suggest that when presenting graphs that contain an underlying long-term trend and substantial short-term variability, spontaneous interpretation of the long-term trend is not guaranteed – indeed fewer than half the participants in the study described any kind of trend. Of the participants who did not describe a trend, the majority did describe short-term variability. Conversely, few of those who described a trend mentioned short-term variability. Hence, other than describing what the data in the graph represented (snow cover over time) which corresponded with the graph title, participants typically only described one aspect of the data (either trend or variability). It's possible that participants felt that they had to only describe the most salient aspect, rather than all aspects – i.e. it might be that they did encode the other characteristic, but did not mention it. Conversely it might be that participants only encoded the characteristic of the data that they described.

Although these two possibilities are not differentiated in this pilot study, the results are consistent with studies that have found impaired task performance with line graphs that contain a high level of variability compared with graphs with less variability (Correll, Albers, Franconeri, & Gleicher, 2012; Carswell, Emery, & Lonon, 1993). Indeed, the ratio between the strength of the trend (i.e. the angle) and the extent of the short-term variability (i.e. the vertical spread) may be

important in determining to what extent different characteristics are salient, to what extent spatial processing is required, and to what extent task performance is impaired. The ratio between the strength of a trend and the extent of short-term variability can be conceptualised as signal (trend) and noise (variability). Snow cover (as plotted in the stimuli graph in this pilot) has a comparatively low signal-to-noise ratio (Krasting, Broccoli, Dixon & Lanzante, 2013). When data that has a greater signal-to-noise ratio are plotted in a line graph, the connected line may contain fewer visual chunks (by virtue of less short-term variability relative to the strength of the trend) and/or may be encoded as a single line, making the identification of a trend easier.

While this pilot study does not identify the extent to which participants mentally encode and process trend information relative to variability information, the data do suggest that when asked to describe a complex time-series, trend information may not be salient. Mental representations, as opposed to verbal descriptions, of long-term trends and short-term variability are considered next in Study 5.

Study 5: Can language support spatial inferences for trends?

The pilot data from Experiment 1 indicate that the long-term trend may not be readily interpreted in graphs that also show short-term variability. The purpose of Study 5 was to investigate whether people encode representations of trends and short-term variability when looking at complex time-series graphs. Furthermore, finding in Study 4 that trends may not be readily interpreted, Study 5 also asked whether language can support the identification of long-term trends.

Language can provide user-goals, which are thought to activate relevant schema and guide visual-spatial attention (Brunyé & Taylor, 2009; Rothkopf, Ballard, & Hayhoe, 2007; Yarbus, 1967). Attending to spatial language when encoding visual scenes can support spatial reasoning (Loewenstein & Gentner, 2005), influence memory for a scene (Feist & Gentner, 2007), and affect the degree to which static images are mentally animated (Coventry, et al., 2013). Therefore, using language to convey the importance of the long-term trend might direct attention to visual features that support encoding of the trend and influence cognition about the trend. Furthermore, presenting this as a ‘warning’ can make the information salient and increase the likelihood that it is acted upon (Wogalter, et al., 1987).

However, linguistic information, including warnings, can be ignored by individuals (Eiriksdottir & Catrambone, 2011; Wogalter, et al., 1987). Even if read, linguistic information may be shallowly processed (LeFevre & Dixon, 1986). Further, although language might be intrinsically tied to flexible spatial skills (Hermer-Vazquez, Spelke & Katsnelson, 1999), individuals can instead rely on visual cues, weighted by prior experience, to support spatial processing (Learmonth, Newcombe, Sheridan, & Jones, 2008). Hence, a linguistic warning might not support interpretation of trends from time-series graphs.

The aim of Study 5 was therefore to test whether a linguistic warning that provides a strategy for interpreting long-term trends (by ignoring task-irrelevant features) would improve encoding of the long-term trend. Furthermore, if a warning is effective, the study asks to what extent the warning is long-lasting, and

whether the effect is driven by changes in visual attention (measured using eye tracking) or whether the warning might merely provide a schema to help organize visual information into long-term memory, without affecting visual attention directly. Informed by previous work (Brunyé & Taylor, 2009; Peebles & Cheng, 2003), it was predicted that the warning would direct visual attention to information consistent with a mentally superimposed line of best of fit.

The study also manipulated a perceptual feature of time-series graphs – the number of intermediary x-axis tick marks and labels. In line with evidence of interactions between top-down and bottom-up processes (Hegarty, Canham, & Fabrikant, 2010), it was hypothesized that intermediary x-axis tick marks and labels might provide salient cues that direct attention to short-term changes in the data, resulting in poorer spatial representation of the long-term trend.

Method

Design

To test spatial representations of the long-term trend and short-term variability, a forced choice task was employed in which participants were shown a graph to study and were then asked to make a ‘same’ or ‘different’ judgment on a following test graph. The test graph was either identical to the study graph (same); had the same overall pattern as the study graph but with a different gradient (gradient different); had the same gradient as the study graph but with exaggerated peaks and troughs (amplitude different); or was completely different to the study graph (completely different). The number of x-axis ticks, either 2, 5 or 9, was varied across each type of test graph (see Figure 13 for examples).

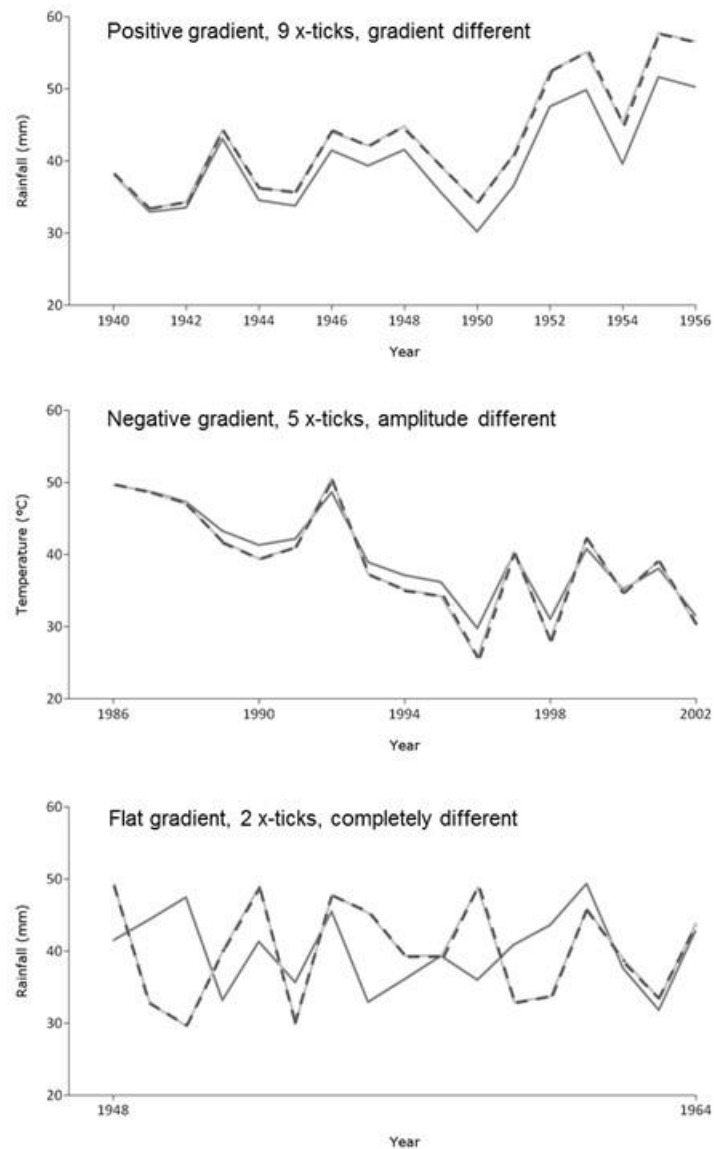


Figure 13: Three examples of study graphs (solid line) and associated test graphs (dashed line) shown here together. Study and test graphs both used solid lines for stimuli presentation and were shown sequentially in the experiment.

To test the effect of a linguistic warning on cognition of the graph, participants were randomly allocated to either receive a warning at the start of the study, or to receive no such warning. The warning read:

“WARNING When looking at graphs, people are often misled by extreme data points – short-term fluctuations in the data can obscure the long-term trend. To avoid errors, it is useful to ignore extreme data points to correctly identify the long-term trend.”

The experiment therefore employed a 4 (Test Graph) x 3 (X-ticks) x 2 (Warning) design, with test graph and x-ticks as within participant variables and warning as a between participant variable.

Participants

Forty undergraduate students (29 female, 11 male) from the University of East Anglia took part in the study in return for course credit or a nominal payment. Average age was 21 years (range 18-30 years). Sample size was informed via power analysis to detect a medium effect size ($\eta p^2 = .060$).

Apparatus

A Tobii TX300 Eye Tracker (Tobii Technology AB, Danderyd, Sweden) with integrated TFT LCD monitor (51cm x 29cm) set to 1280 x 720 pixels was used for stimulus presentation and collection of eye gaze data at 300Hz. Eprime Version 2.0 (Psychology Software Tools Inc., Sharpsburg, USA) was used to control stimulus presentation and record data. Responses for same-different trials were mapped to the ‘Z’ and ‘M’ keyboard keys, which were reversed and counterbalanced between warning conditions. Verbal responses were recorded via a headset microphone. Eye gaze data were analyzed using OGAMA Version 4.5 (A. Voßkühler, Freie Universität Berlin, Germany), using default parameters for fixation detection.

Linguistic warning

The linguistic warning was displayed in 28pt Calibri, center aligned.

Graph stimuli – ‘same-different’ trials

Study time-series graphs were created (1126 x 510 pixels), each plotting 17 data points. Twelve initial datasets were created for the study graphs for ‘same-different’ trials, four of which showed an underlying positive long-term trend, four a negative long-term trend and four a flat long-term trend (Figure 13). Data points for each graph were created by sampling residuals at random from a normal distribution, which were then applied to a baseline positive (gradient = 1.0, intercept = 30), negative (gradient = -1.0, intercept = 50) or flat (gradient = 0.0, intercept = 40) linear trend graph. The x-axis was labeled ‘Years’ and the y-axis was labeled either as “Medication use (doses)”, “Infections (patients)”, “Temperature (°C)”, “Rainfall (mm)”, “Income (GBP £)”, or “Expenditure (USD \$)”. The x-axis covered a range of 16 years, with the starting year always between 1900 and 1994. The y-axis covered a range of 40 units, starting at 20 and finishing at 60 units. A caption was created for each graph that simply read “[variable] over time”. Three study graphs – one with a positive trend, one with a negative trend, and one with a flat trend – were allocated to each of the four test graph conditions (same, gradient different, amplitude different, completely different).

For each of the twelve study graphs, a corresponding test graph was then created. For the three study graphs allocated to the ‘same’ condition, test graphs were identical to the study graph. Test graphs for the three study graphs allocated to the ‘gradient different’ condition had a subtly different gradient to the study graph (transformation of the y values of the study graph: $y' = y \pm 0.4x$). The direction of the transformation, i.e. shift upward applying $+0.4x$, or a shift downward applying $-0.4x$, was matched to the gradient of the line of best fit for the study graph. Flat trend graphs had gradients close to, but not exactly equal to 0, owing to the random sampling of residuals. Therefore, positive long-term trend study graphs had test graph pairings that became steeper (more positive), negative long-term trend study graphs had test graph pairings that also became steeper (more negative), and flat long-term trend study graphs with a line of best fit gradient > 0 had test graph pairings that became more positive and flat long-term

trend study graphs with a line of best fit gradient < 0 had test graph pairings that became more negative.

Test graphs for the three study graphs allocated to the amplitude different condition had extended peaks and troughs compared to the study graph (residuals multiplied by 1.4). For the three study graphs allocated to the ‘completely different’ condition, three new graphs were produced to serve as test graphs.

For each of the 12 study-test graph pairings, three variants were then created, showing 2, 5 and 9 x-ticks (Figure 13), resulting in a total of 36 study-test graph pairings.

Graph stimuli – ‘describe’ filler trials

A further group of graphs was created (using the same pixel dimensions, plotting the same number of data points, and using the same labelling as for the same-different trials), which acted as filler trials on which participants were tasked to describe the graph. Three initial datasets were created, one with a positive long-term trend, one with a negative long-term trend, and one with a flat long-term trend. For each of these initial datasets, three graph variants were then created, showing 2, 5 and 9 x-ticks, resulting in a total of 9 graphs for the ‘describe’ filler trials.

Graph stimuli – ‘comprehension’ filler trials

A final group of study time series graphs was created (using the same pixel dimensions, plotting the same number of data points, and using the same labelling as for the same-different trials), which acted as filler trials on which participants were asked to answer a comprehension question about the graph. Nine initial datasets were created, three with a positive long-term trend, three with a negative long-term trend, and three with a flat long-term trend. In this instance, within each set of positive, negative and flat graphs, one graph showed 2 x-ticks, one showed

5 x-ticks and one showed 9 x-ticks. In total there were 9 graphs for ‘true-false comprehension’ filler trials.

Areas of interest (AOI)

AOIs were defined for each study graph by first determining a circle around each data point with a maximum diameter that would avoid overlapping adjacent AOIs (58 pixels), i.e. the largest mutually exclusive area that could be defined for a data point radiating from the centre of each data point. A parallelogram (2.0 x 34.5 degrees of visual angle) was then fitted over the line of best fit of the plotted data, determined by linear least squares regression. The height of the parallelogram was the same size as that used for the data points (58 pixels), and the length of a parallelogram was determined by the distance between the outer edges of the first and last data point AOI (1002 pixels). The parallelogram formed the line of best fit AOI (6.3% of screen area). A convex hull was then determined around the outer edges of the defined shapes, which formed the whole data AOI (mean 22.1% of screen area). An extreme data AOI was defined as the area of the whole data AOI that sat outside of the line of best fit AOI (mean 15.8% of screen area) (Figure 14).

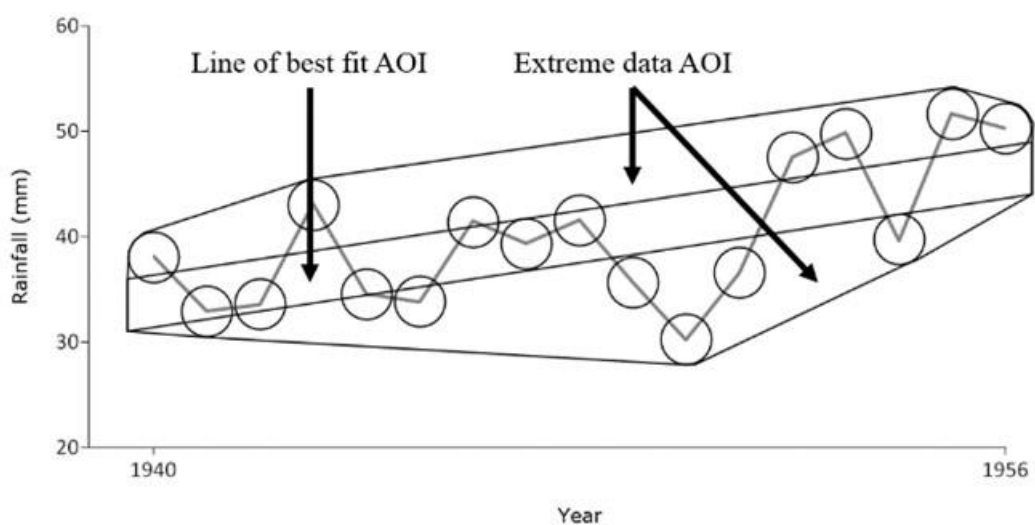


Figure 14: Line of best-fit AOI and extreme data AOI for one of the 24 study graphs.

Procedure

Participants were informed that the study was investigating how people understand line graphs and they then received instructions on screen before a practice block of trials. The eye tracker was then calibrated. Participants were randomly allocated to either the warning or no warning condition, with the requirement of two equal sized groups (20 participants in each group). Participants in the warning condition then received the warning on screen and were instructed to read it before starting the first of three blocks of trials. Participants in the no warning condition simply started the first block of trials after eye tracker calibration. Each trial consisted of a study phase (Figure 15) during which participants were asked to look at and study the caption and the graph. The caption was presented prior to the graph to help control time spent reading the caption. The test phase began by indicating which task would follow, i.e. same-different, true-false, or describe (true-false and describe tasks were included to encourage participants to study the graphs in a naturalistic way and to ensure depth of encoding). For same-different trials, participants then made a same-different judgment about a test caption and then a same-different judgment about a test graph (i.e. comparing to their memory for the study caption and study graph). Participants were instructed to give a response as quickly as possible when the test caption/graph appeared.

Trials were presented in three blocks. Each block contained 18 trials – 12 same-different trials, three true-false filler trials and three describe filler trials – presented in random order. Within a block, each of the initial 12 same-different study datasets appeared once, with each x-tick variant appearing in separate blocks (i.e. a same-different study graph dataset only appeared once in a block). Study-test graph pairings were allocated to blocks such that each block contained three ‘same’ trials, three ‘amplitude different’ trials, three ‘gradient different’ trials and three ‘completely different’ trials. Furthermore, each block contained four positive trend same-different study graphs, four negative trend same-different study graphs, and four flat trend same-different study graphs. In addition, each block contained four study-test graphs for each of 2, 5 and 9 x-ticks. Hence, for same-different trials, trial type, trend and x-ticks was balanced in

each block. Each block also balanced trend and x-ticks among ‘describe’ and ‘comprehension’ filler trials. See Appendix 2 Table A2-1 and Table A2-2 for full allocation of trials to blocks.

The specific trials allocated to a block was identical for all participants, but the order in which trials appeared within a block was randomised for each participant. Further, the order of the blocks was counterbalanced across participants. The eye tracker was re-calibrated at the start of each block. At the end of the third block, participants in the warning condition were asked what they remembered about the warning. The study lasted approximately 1 hour.

Results

Data screening

Due to the importance of encoding the warning, a strict exclusion criterion was used, requiring accurate recall of the warning at the end of the study. Only same-different trials where participants correctly remembered the caption and then went on to make a judgement about the graph were included in the analyses. Six participants were removed from further analyses: four participants in the warning condition who could not recall the warning at the end of the study; one participant who subsequently reported monocular vision impairment; and one participant whose accuracy on completely different trials was 11% (lower than three standard deviations from mean accuracy). Following data screening, 34 participants were included in data analysis, 18 in the no warning condition and 16 in the warning condition.

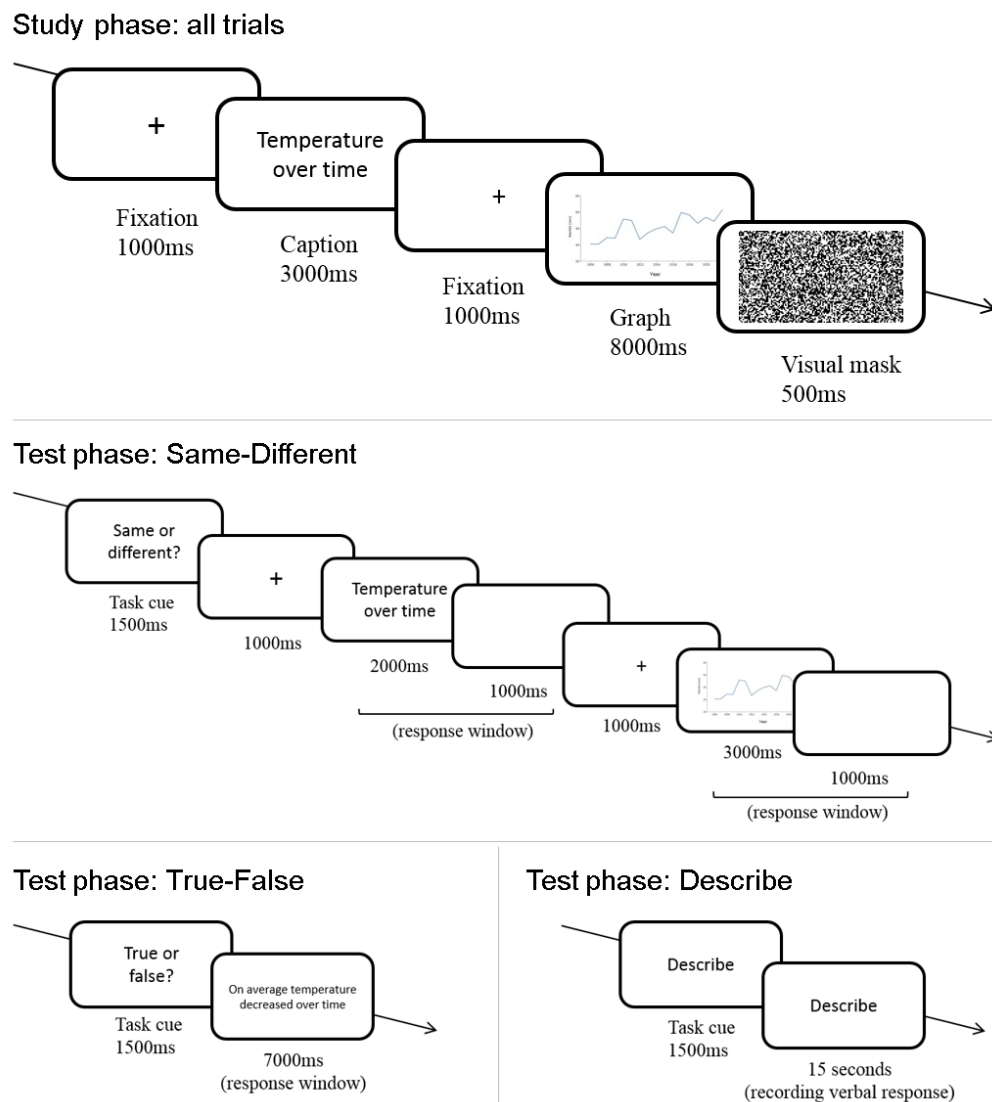


Figure 15: Presentation of same-different and filler trials.

Task performance.

Sensitivity to detect differences between the graphs on same-different trials was measured using d' , calculated using the log-linear rule (Hautus, 1995). There was no significant difference between the warning and no warning groups on ability to discriminate between completely different trials, $t(32) = -0.341$, $p = .735$, $d = 0.117$, 95% CI [-0.558, 0.790]. To assess sensitivity to detect subtle changes between study and test graphs, participants' d' scores for amplitude and gradient

sensitivity were analyzed with a 2 (Test Graph [amplitude different, gradient different]) x 3 (X-ticks [2, 5, 9]) x 2 (Warning [no warning, warning]) mixed ANOVA.

Means and standard deviations for each cell of the analysis are provided in Appendix 3, Table A3-1. There was no main effect of test graph, x-ticks, or warning (Table 5). However there was a significant interaction between test graph and warning, $F(1,32) = 4.399, p = .044, \eta p^2 = .121$ (Figure 16). Participants in the no warning condition performed significantly worse on gradient different trials than amplitude different trials: $t(17) = -3.381, p = .004, d = -0.823, 95\% \text{ CI} [-1.364, -0.263]$; whereas those in the warning condition performed about equally on gradient different trials and amplitude different trials: $t(15) = 0.112, p = .912, d = 0.030, [-0.497, 0.556]$. There were no other significant two-way interactions and no three-way interaction (Table 5).

Table 5. Study 5 mixed ANOVA table (test graph x x-ticks x warning); * indicates significance at the .05 level.

Source	Test	<i>p</i> -value	ηp^2
Main effects			
test graph	$F(1,32) = 3.655$.065	.103
x-ticks	$F(2,64) = 0.365$.696	.011
warning	$F(1,32) = 0.034$.855	.001
Two-way interactions			
test graph x warning	$F(1,32) = 4.399$.044*	.121
test graph x x-ticks	$F(2,64) = 0.060$.942	.002
x-ticks x warning	$F(2,64) = 2.512$.089	.073
Three-way interaction			
test graph x x-ticks x warning	$F(2,64) = 0.273$.762	.008

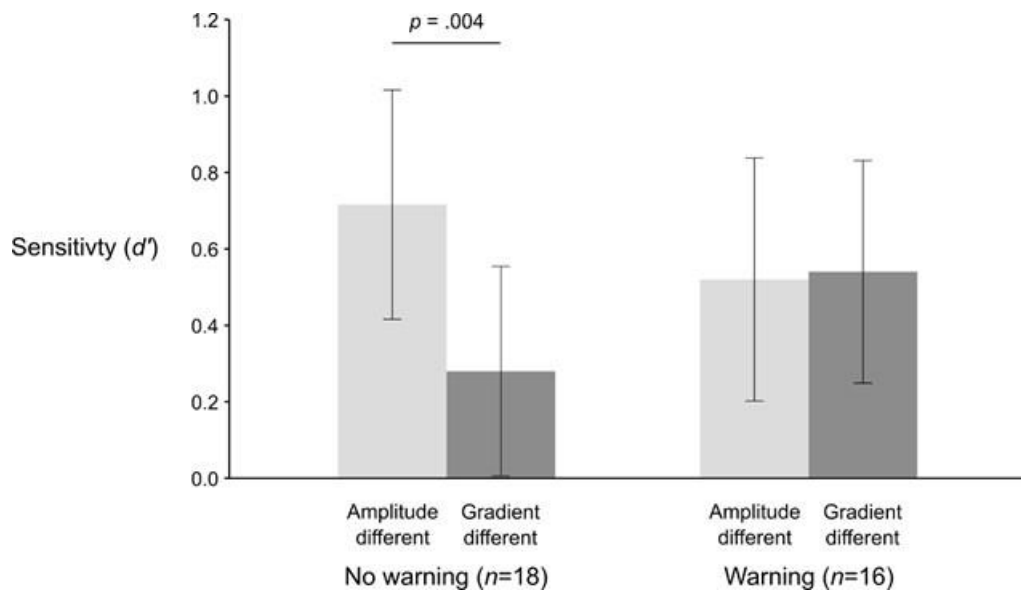


Figure 16: Average sensitivity (d') for amplitude different and gradient different trials in each group, with 95% confidence intervals.

To investigate if the effect of the warning on gradient performance deteriorated over time, d' values were recalculated by collapsing data across x-ticks (as there was no significant x-ticks main effect or interaction), and then splitting the data by block of trials, i.e. first block, intermediary block, last block. A 2 (Test Graph) x 3 (Block) x 2 (Warning) mixed ANOVA was then performed. Means and standard deviations for each cell of the analysis are provided in Appendix 3, Table A3-2. Results were consistent with the first mixed ANOVA (i.e. a significant Test Graph x Warning interaction), but there was no three-way interaction between test graph, warning and block (Table 6).

Table 6. Study 5 mixed ANOVA table (test graph x block x warning); * indicates significance at the .05 level.

Source	Test	<i>p</i> -value	ηp^2
Main effects			
test graph	$F(1,32) = 4.092$.051	.113
block	$F(2,64) = 1.116$.334	.034
warning	$F(1,32) = 0.014$.906	<.001
Two-way interactions			
test graph x warning	$F(1,32) = 4.319$.046*	.119
test graph x block	$F(2,64) = 0.509$.603	.016
block x warning	$F(2,64) = 0.330$.720	.010
Three-way interaction			
test graph x block x warning	$F(2,64) = 0.026$.974	.001

Visual attention

To investigate if the improved discriminability of the gradient found in the warning condition might be driven by differences in visual attention during encoding, fixation durations for the AOIs of the study graphs were calculated. Four participants were excluded from further analysis as they had poor eye tracking calibrations (two participants from each of the warning conditions, leaving 16 participants in the no warning group and 14 participants in the warning group). Same-different trials in which a correct response was given to the caption and a response was given to the test graph, all trials for the true-false task in which a response was given, and all trials for the describe task were included in the analysis. However, individual trials were excluded if >15% of eye tracking samples were missing, or if there was a continuous period >700ms of data missing (10.7% of trials). As there was no main effect or interaction of x-ticks in the d' data, fixation data were collapsed across x-ticks.

The data were checked to see if the warning influenced the total fixation duration for the whole data area compared to the no warning group, finding no significant difference: $t(28) = 1.288, p = .208$ (two-tailed, equal variances assumed), $d = 0.471, 95\% \text{ CI } [-0.261, 1.195]$. Fixation durations for the line of best fit AOI and extreme data AOI were then compared. Homogeneity of variances between the warning and no warning groups could not be assumed for total fixation data for the line of best fit AOI or the extreme data AOI; Levene's test for equality of variances were, $F(1,28) = 9.121, p = .005$; and $F(1,28) = 5.285, p = .029$, respectively. Therefore, separate independent t -tests were performed on the data in line with a priori predictions.

Participants in the warning condition spent significantly longer fixating on the line of best fit area than participants who did not receive the warning, $t(19.802) = 2.119, p = .024$ (one-tailed, equal variances not assumed), $d = .804, 95\% \text{ CI } [0.050, 1.545]$. Conversely, there was no significant difference for the extreme data area, $t(25.137) = -0.352, p = .728$ (two-tailed, equal variances not assumed), $d = -0.125, [-0.842, 0.594]$ (see Table 7 for fixation durations).

Table 7: Study 5 mean (M) and standard deviations (SD) of fixation duration in ms during study for each AOI.

Area of interest	No warning ($n = 16$)		Warning ($n = 14$)	
	M	SD	M	SD
Line of best fit	1426	(432)	1919	(772)
Extreme data	1587	(586)	1525	(356)
Whole data	3013	(884)	3444	(952)

Discussion

Compared to the no warning control group, the linguistic warning improved spatial representations of long-term trends relative to representations of short-term variability. The effect of the warning lasted for the duration of the study (~50

minutes) and did not impair representations of extreme data points. Eye-tracking data were consistent with task performance, and indicated that the warning acted directly on visual attention by increasing attention towards the line of best fit, but not drawing attention away from the extreme data points. There was no influence of the number of x-axis ticks/labels; they may not have been sufficiently salient to guide attention to the extreme data (cf. Fabrikant, Hespanha, & Hegarty, 2010).

These findings are consistent with prior work indicating that goals, instantiated using language, direct attention to goal-relevant information (Brunyé and Taylor, 2009). However, the effect of the warning may have directed attention to information congruent with the goal of the instruction (i.e. line of best fit), suggesting that warnings are goal-specific. Alternatively, the warning might have simply increased vigilance and attention more generally, suggesting that warnings can be goal-independent. In support of this latter view, Study 5 did not see a drop in performance for the amplitude trials when the warning asked to focus on trend, even though the warning instructed people to ignore extreme data.

Study 6: What language supports spatial inferences for trends?

To differentiate between the goal-specific and goal-independent accounts of the effect of the linguistic warning found in Study 5, in Study 6 the goal of the warning was manipulated to either encourage encoding of the trend (i.e. ignore extreme data and identify trend), or to encourage encoding of extreme data (i.e. ignore trend and identify the extreme data). If a warning is goal-specific it was expected that warnings would support representations of goal-congruent information but not support representations of goal-incongruent information. Conversely, if a warning supports representations of goal-congruent and goal-incongruent information, this would suggest warnings are goal-independent.

However, an alternative possibility is that the inclusion of an instruction to ignore something (i.e. either extreme data or trend) might paradoxically increase attention to that information rather than diminish attention to it. Instructions not to think of something, e.g. to ignore something, or to try to suppress a thought can increase the frequency of the thought (Abramowitz, Tolin, & Street, 2001). Furthermore, representations of stimuli that are to be ignored may still be encoded into memory and subsequently influence behaviour (Grison, Paul, Kessler, & Tipper, 2005; Grison, Tipper, & Hewitt, 2005). Being told to ignore information might act in the same way as an explicit instruction to attend to that information. Therefore, a warning to ignore one aspect and identify another aspect, might in fact serve as a warning to attend to both aspects. In Study 6, the content of the warning was therefore manipulated, providing either an 'ignore and identify' instruction, or only an 'ignore' instruction. A control 'no warning' group was also included to check for replication of the effect of the warning found in Study 5 (i.e. comparing 'no warning' to the warning to 'identify trend and ignore extreme data points').

Method

Design

The same forced-choice task as Study 5 was used to test spatial representations of the long-term trend and short-term variability, using the same four test graph conditions (same, gradient different, amplitude different, completely different). To test the effects of warning goal and informational content on cognition of the graphs, participants were randomly allocated to either receive one of four warnings, or to receive no warning (control group). The four warnings were:

1. Trend goal, ignore and identify: “WARNING When looking at graphs, people are often misled by extreme data points - short term patterns in the data can obscure the long term trend. To avoid errors, it is useful to ignore extreme data points to correctly identify the long term trend.”
2. Ignore extreme only: “WARNING When looking at graphs, people are often misled by extreme data points. To avoid errors, it is useful to ignore extreme data points.”
3. Extreme goal, ignore and identify: “WARNING When looking at graphs, people are often misled by long term trends - long term patterns in the data can obscure the extreme data points. To avoid errors, it is useful to ignore the long term trend to correctly identify the extreme data points.”
4. Ignore trend only: “WARNING When looking at graphs, people are often misled by long term trends. To avoid errors, it is useful to ignore the long term trend.”

The main experiment was therefore a 4 (Test Graph) x 2 (Warning Goal [trend, extreme data points]) x 2 (Informational Content [ignore and identify, ignore only]) design, with test graph as a within participant variable and warning goal and informational content as between participant variables. The design also incorporated the ‘no warning’ group as a control. As d' values for the temporal analysis in Study 5 were of limited precision (calculated over a maximum of three trials), in Study 6 the number of same-different trials were increased and temporal resolution for precision was sacrificed by considering two time blocks (block 1

and block 2), rather than three, (d' values in the current study were therefore calculated over a maximum of six trials).

Participants

One-hundred and thirty-one undergraduate students (107 female, 24 male) from the University of East Anglia took part in the study in return for course credit or a nominal payment. Average age was 21 years (range 18-55 years). Participants were recruited until there were 18 participants in each group (after accounting for participants removed per data screening criteria).

Apparatus

TFT LCD monitors (51cm x 29cm) set to 1280 x 720 pixels were used for stimulus presentation and Eprime Version 2.0 (Psychology Software Tools Inc., Sharpsburg, USA) was used to control stimulus presentation and record data. Response keys and mappings were the same as Study 5.

Linguistic warnings

The linguistic warnings were displayed in the same font size and style as Study 5.

Graph stimuli – ‘same-different’ trials

Forty-eight initial datasets were created using the same format and procedure as Study 5, 16 of which showed a positive long-term trend, 16 a negative long-term trend and 16 a flat long-term trend. Twelve study graphs – four with a positive trend, four with a negative trend, and four with a flat trend – were allocated to each of the four test graph conditions (same, gradient different, amplitude different, completely different).

Corresponding test graphs for the same and completely different conditions were created as per Study 5. Test graphs for the gradient different

condition were created in the same manner as Study 5, except the transformation of the y values of the study graph was: $y' = y \pm 0.48x$. Test graphs for the amplitude different condition were created by multiplying the residuals of the study graph by a factor of 1.5. Compared to Study 5, the test graphs for the gradient different and amplitude different therefore had slightly larger changes to their study graphs with the aim of improving d' . There were a total of 48 study-test graph pairings.

Graph stimuli – filler trials

A further twenty-four datasets were created using the same format and procedure as Study 5, which acted as filler trials on which participants were asked to answer a comprehension question about the graph. Eight had a positive long-term trend, eight had a negative long-term trend, and eight had a flat long-term trend. In contrast to Study 5, there were no ‘describe’ filler trials in the present study.

Procedure

The procedure was the same as Study 5, except that eye tracking was not employed, each of the three blocks consisted of 16 same-different trials, eight true-false comprehension trials, and there were no describe filler trials. For the 16 same-different trials in a block, there were four trials for each of the different test graph conditions. One block contained six positive trend graphs, five negative trend graphs and five flat trend graphs. A second block contained five positive trend graphs, six negative trend graphs and five flat trend graphs. A third block contained five positive trend graphs, five negative trend graphs and six flat trend graphs (See Appendix 4 for full allocation of graphs to blocks). As per Study 5, the specific trials allocated to a block was identical for all participants, but the order in which trials appeared within a block was randomised for each participant. Further, the order of the blocks was counterbalanced across participants.

Participants were randomly allocated to either the no warning condition, or one of the four warning conditions, with the requirement of five equal sized

groups after data screening (data were screened during the process of data collection). The study lasted approximately 50 minutes.

Results

Data screening

As per Study 5, accurate recall of the warning at the end of the study was required; and only same-different trials in which a correct response was given to the test caption and a response was given to the test graph were included in the analyses. The total number of participants in each warning condition before applying screening criteria, were 24 in the ‘no warning’ condition; 24 in the ‘trend goal, ignore and identify’ condition; 31 in the ‘extreme goal, ignore and identify’ condition; 22 in the ‘ignore extreme only’ condition, and 30 in the ‘ignore trend only’ condition.

Forty-one participants were removed from further analyses: 28 participants who could not correctly remember the warning when asked at the end of the study, and 13 participants whose accuracy on completely different trials was < 31% (lower than three SD from mean accuracy). After applying the screening criteria, eighteen participants remained in each warning condition. As per the methodology used in Study 5, sensitivity to detect differences between the graphs on same-different trials was then measured using d' , calculated using the log-linear rule (Hautus, 1995).

Effect of a warning – comparison with Study 5

To investigate whether the data from the ‘no warning’ and ‘identify trend, ignore extreme’ warning replicated the findings from Study 5, data from these two groups were compared. There was no significant difference between the warning and no warning groups on ability to discriminate between completely different trials, $t(34) = -0.086$, $p = .932$, $d = -0.029$, 95% CI [-0.682, 0.625]. Participants’ d' scores for amplitude and gradient sensitivity were then analyzed with a 2 (Test Graph) x 2 (Block) x 2 (Warning) mixed ANOVA.

Means and standard deviations for each cell of the analysis are provided in Appendix 5, Table A5-1. Consistent with Study 5, there was no main effect of test graph, block, or warning (Table 8). Converse to Study 5, however, the interaction between test graph and warning did not reach statistical significance, $F(1,34) = 2.113$, $p = .155$, $\eta p^2 = .059$. No other two-way interactions reached statistical significance and neither did the three-way interaction. Given the *a priori* prediction of an interaction between test graph and warning (per the results of Study 5), and in line with best practice in considering effect sizes and confidence intervals (Cummings, 2014) the patterns of the data for the test graph by warning interactions in each study are shown in Figure 17.

Table 8. Study 6 mixed ANOVA table (test graph x block x warning); * indicates significance at the .05 level.

Source	Test	<i>p</i> -value	ηp^2
Main effects			
test graph	$F(1,34) = 0.230$.635	.007
block	$F(1,34) = 0.098$.757	.003
warning	$F(1,34) = 0.547$.465	.016
Two-way interactions			
test graph x warning	$F(1,34) = 2.113$.155	.059
test graph x block	$F(1,34) = 1.825$.186	.051
block x warning	$F(1,34) = 1.031$.317	.029
Three-way interaction			
test graph x block x warning	$F(1,34) = 0.452$.506	.013

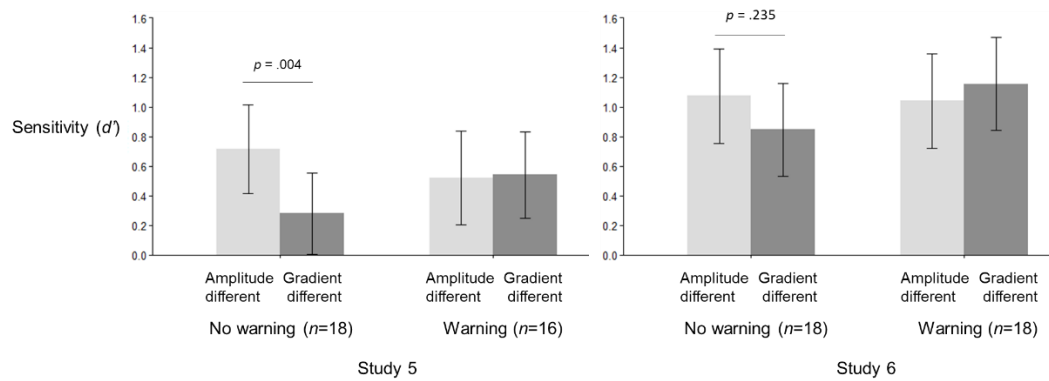
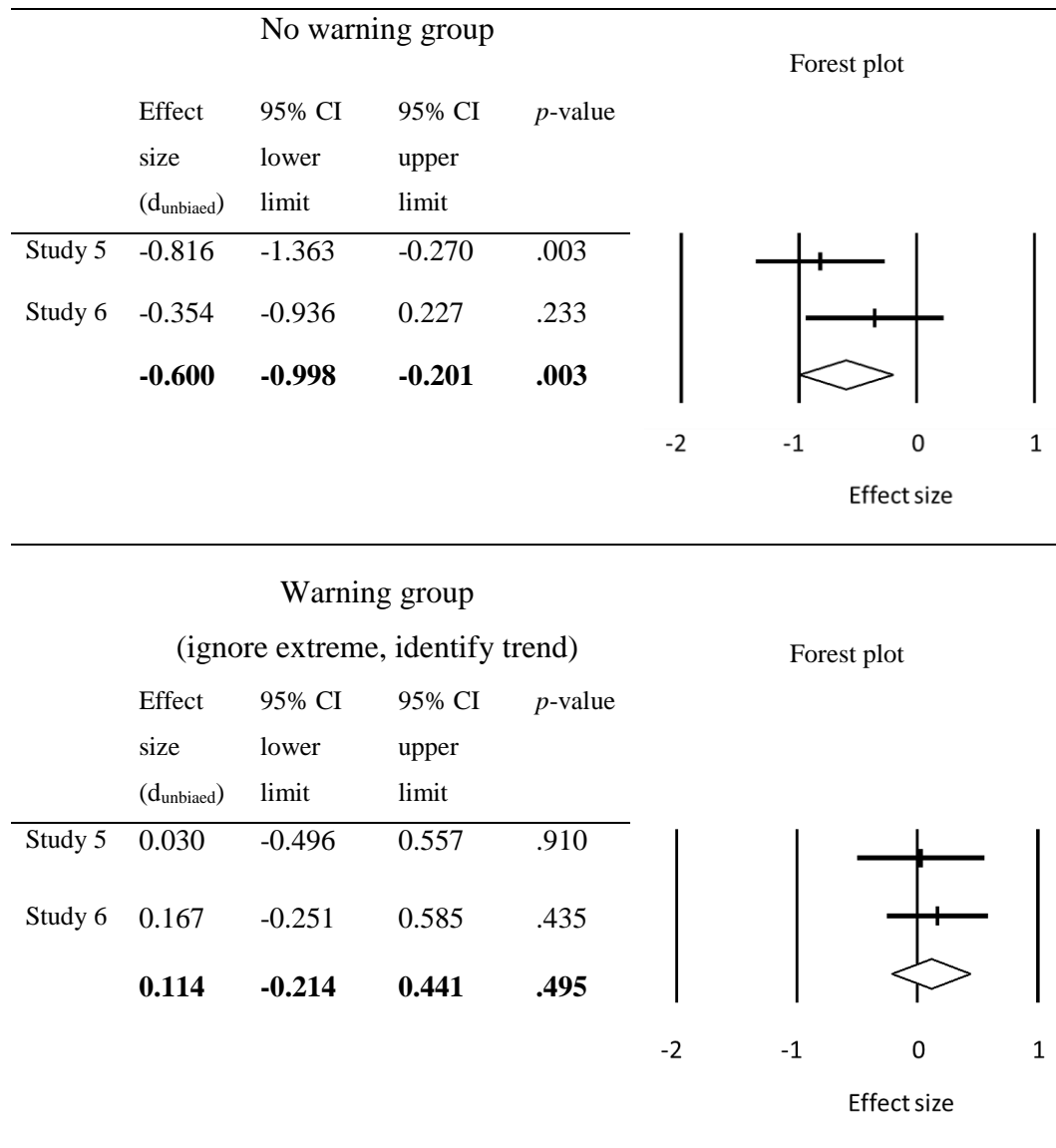


Figure 17. Comparison of interaction between test graph and warning in Study 5 (left) and Study 6 (right). Error bars represent 95% confidence intervals. Note: d' values are generally greater in Study 6 than Study 5 as expected as stimuli were adapted to in Study 6 to reduce the potential for floor effects. The pattern of differences between conditions is consistent across both experiments.

To evaluate differences in sensitivity between gradient different trials and amplitude different trials in the no warning and warning conditions across evidence collected in Study 5 and 6, a meta-analysis using a fixed-effects model was conducted for the no warning and warning groups (Table 9). Analysis indicated a reliable difference in no warning groups, where sensitivity to detect differences was better for amplitude different trials than gradient different trials ($p = .003$). Conversely, no such difference was found in warning groups ($p = .495$).

Table 9. Meta analyses of the effect size of the paired difference between sensitivity for gradient differences and amplitude differences in Study 5 and 6; negative effect sizes indicate better performance on amplitude different trials, positive effect sizes indicate better performance on gradient different trials.



Effect of user goal and informational content.

To investigate the effect of user goal and informational content on sensitivity to detect differences, analysis of data across the four warning conditions was conducted. There was no significant difference between any of the four warning groups on sensitivity (d') on completely different trials, in block 1 or block 2: $F(3,67) = 0.117, p = .950, \eta p^2 = .005$; $F(3,68) = 0.484, p = .694, \eta p^2 = .021$,

respectively. Means and standard deviations for each of the warnings are provided in Appendix 5, Table A5-2.

Performance across conditions was then compared by submitting d' scores to a 2 (Test Graph [gradient different, amplitude different]) x 2 (Warning User Goal [long-term trend, extreme data]) x 2 (Informational Content [ignore and identify, ignore only]), mixed ANOVA. Means and standard deviations for each cell of the analysis are provided in Appendix 5, Table A5-3. There was no main effect of test graph, user goal, or informational content (Table 10). There was a significant two-way interaction between test graph and informational content, $F(1,68) = 5.140$, $p = .027$, partial $\eta p^2 = .070$. Post-hoc tests did not identify reliable differences between means, but the interaction was likely driven by greater sensitivity to gradient different trials in the 'ignore and identify' warnings compared to 'ignore' warnings; $t(70) = 1.905$, $p = .061$ (two-tailed), $d = 0.450$, 95% CI [-0.020, 0.916] (Figure 18). No other interactions were significant at the .05 level (see Table 10).¹

¹ Data were collapsed over blocks, and block was not included in the reported mixed ANOVA as no previous analyses found a main effect or interaction involving block. However, when including block in the ANOVA results were consistent with those reported above, i.e. a significant test graph x informational content interaction, and no other significant interactions or main effects.

Table 10. Study 6 mixed ANOVA table (test graph x warning user goal x warning informational content); * indicates significance at the .05 level.

Source	Test	<i>p</i>	ηp^2
Main effects			
test graph	$F(1,68) = 0.134$.716	.002
warning user goal	$F(1,68) = 1.646$.204	.024
warning informational content	$F(1,68) = 0.831$.365	.012
Two-way interactions			
test graph x warning user goal	$F(1,68) = 0.387$.536	.006
test graph x warning informational content	$F(1,68) = 5.140$.027*	.070
Warning user goal x warning informational content	$F(1,68) = 3.462$.067	.048
Three-way interaction			
test graph x warning user goal x warning informational content	$F(1,68) = 0.530$.469	.008

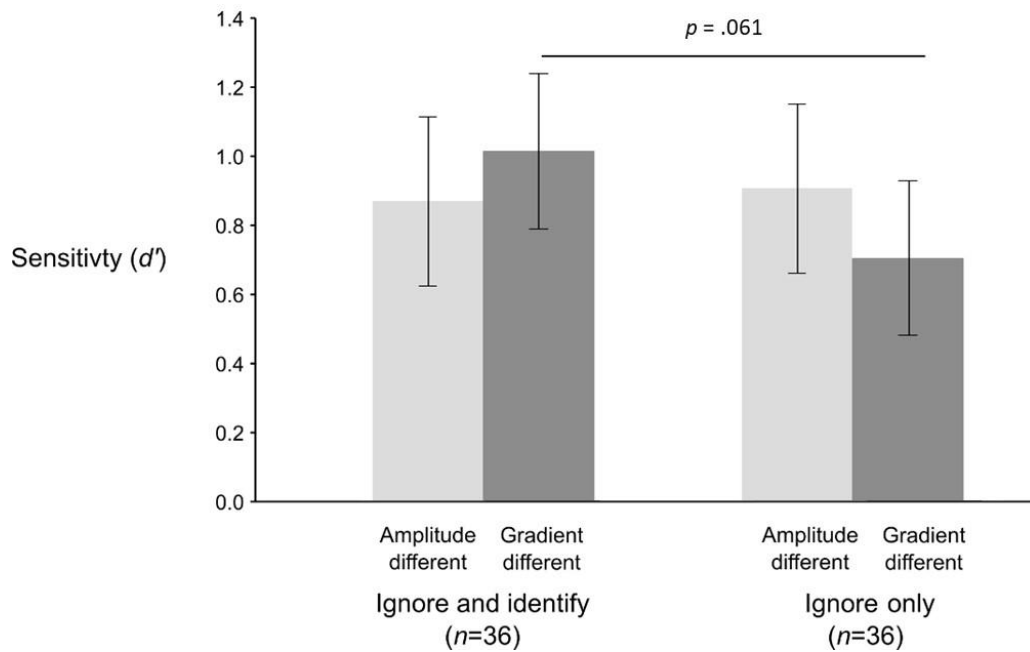


Figure 18. Average sensitivity (d') for amplitude and gradient trials by informational content in Study 6, with 95% confidence intervals. Error bars represent 95% confidence intervals.

To inform the nature of the interaction between informational content and test graph relative to the control (no warning) group, d' scores were submitted to a 2 (Test Graph [gradient different, amplitude different]) x 3 (Informational Content [ignore and identify, ignore only, no warning]), mixed ANOVA (Table 11). Means and standard deviations for each cell of the analysis are provided in Appendix 5, Table A5-4. There was no main effect of test graph; or of informational content. The two-way interaction between test graph and informational content, did not achieve statistical significance.

Table 11. Study 6 mixed ANOVA table (test graph x warning informational content); * indicates significance at the .05 level.

Source	Test	<i>p</i>	ηp^2
Main effects			
test graph	$F(1,87) = 1.665$.200	.019
informational content	$F(2,87) = 0.844$.433	.019
Two-way interactions			
test graph x warning informational content	$F(2,87) = 3.094$.050	.066

Post-hoc analyses to explore the non-significant interaction between test graph x warning informational content were conducted to understand potential relationships². No significant differences were identified between either of the warning groups and the control group (Figure 19).

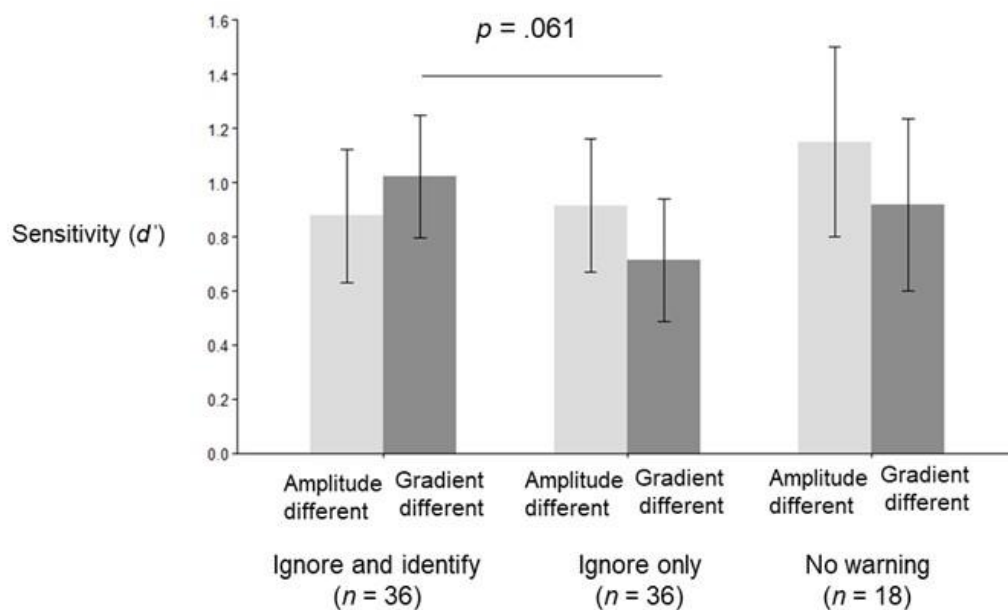


Figure 19. Interaction between test graph and informational content in Study 6, including the no warning group. Error bars represent 95% confidence intervals.

² Per interpretation of *p* values advocated by Fisher (1955).

Discussion

Consistent with Study 5, spatial representations of long-term trends were greater when participants received a linguistic warning instructing them to ignore extreme data and identify trends, compared to a control ‘no warning’ group. Although this difference did not reach statistical significance at the .05 level in Study 6, meta-analysis across Study 5 and Study 6 strongly indicates an effect of the warning.

There is some evidence to suggest that spatial representations of long-term trends were improved when the linguistic warning more generally contained an ‘ignore and identify’ instruction, compared to an ‘ignore only’ instruction, i.e. independent of the specific warning goal. However, there was no evidence for (paradoxical) effects of thought suppression when the ‘ignore’ warnings were presented in isolation – i.e. a warning to ignore either trends or extreme data points did not paradoxically support spatial representations. In the context of identifying long-term trends, the data are broadly consistent with the goal-independent account – a warning to identify something in a graph and to ignore something else *increases* vigilance and attention, which then supports representations of long-term trends. However, this effect was only found in comparison to the ‘ignore only’ warnings and not the control ‘no warning’ group. This raises that possibility that a warning to simply ignore something *decreases* vigilance and attention, relative to receiving no warning. Further investigation with improved statistical power is needed to distinguish whether this is indeed the case.

Across the warning goal and informational content manipulations, there was no effect on representations of extreme data points. This suggests that unlike trends, linguistic warnings are not effective at supporting representations of extreme data.

General discussion

Over two experiments, evidence clearly indicates that a goal-congruent linguistic warning can support spatial representations of long-term trends. Furthermore, there is some evidence to suggest that linguistic warnings might confer a benefit, regardless of whether they instruct individuals to identify goal-congruent or goal-incongruent information, specifically in comparison to a warning to ignore something. Importantly we found that instruction to only ‘ignore the long-term trend’ and ‘ignore extreme data points’ did not facilitate improved representations of long-term trends or extreme data points, respectively. Hence, evidence discounts the possibility that suppressing goal-incongruent information might paradoxically facilitate representations of goal-incongruent information from intrusive thoughts.

The findings are therefore broadly consistent with a goal-independent account, in which an instruction to identify something in a visual-display and discount something else, increases attention and vigilance in general, resulting in a more comprehensive representation of the display. The observed differences in visual attention in Experiment 1 lend further support that the linguistic warning acted directly on attentional processes, and are consistent with work that has identified differences in visual attention during encoding following goal-direction instruction when studying maps (Brunyé & Taylor, 2009), and when making spatial inferences from weather charts (Fabrikant, Hespanha, & Hegarty, 2010). However, the current research shows for the first time that a single instantiation of a succinct linguistic warning can quickly influence the nature of mental representations of spatial inferences drawn from complex graphs.

However, it is important to note that in addition to a goal-independent effect, providing a specific user-goal might enable attention to be better targeted. Across the two studies a goal congruent warning improved spatial representations of trends relative to the control group, whereas evidence for the goal-independent effect was only found relative to the ‘ignore’ warnings. Further investigation is needed to differentiate the extent to which attention is facilitated separately by goal-dependent and goal-independent components, and whether a warning to

simply ignore something decreases attention and vigilance. The current work, suggests that such effects may be small-to-moderate and require greater statistical power in follow-up studies.

An effect of linguistic warnings was only found on spatial representations of trends, and not on representations of extreme data. If goal-independent warnings can increase attentional processing in general, then why weren't both aspects of the visual display encoded more accurately? It is suggested that the relative ease in identifying extreme data points (perceptually salient trend reversals) compared to long-term trends (requiring spatial integration of multiple visual chunks) might account for this distinction. Top-down attentional gains afforded by a warning might provide limited or no additional benefit when encoding perceptually salient features, as bottom-up processing is already adapted to this task (Itti & Koch, 2001). In contrast, top-down attentional gains might add valuable visual information that support encoding of features that require effortful cognition to infer.

Research on the comprehension of visual displays has highlighted the role of 'offloading cognition on perception' to aid understanding (e.g. Hegarty, 2011, pp. 452; Shah, Mayer, & Hegarty, 1999;). As an alternative to a linguistic warning, one could plot a trend line directly in the display or employ smoothing techniques to the data to support spatial representations of trends. However, non-experts may perceive scientific graphs as literal descriptions of reality, and have feelings of unease when made aware of such statistical transformations and graphical techniques (Walsh, 2014; Walsh, 2015). Therefore, there might be advantages to using language to guide non-experts to make appropriate spatial inferences themselves, e.g. to improve trust in the data. Further research on these issues will be important to help understand cognition for visualizations of complex real-world datasets, especially in domains where inferences drawn from data are contested. Insights from such research would also help to respond to the call to scale-up cognitive models of visual-spatial display comprehension (Hegarty, 2011).

In conclusion, providing a linguistic warning that includes a goal to attend to something in a visual display, may support improved spatial encoding of the visual display in general (i.e. encoding of features related and unrelated to the goal). These encoding gains may be especially relevant to inferring patterns in the visual display that are not perceptually salient.

Chapter 4: Forming expectations from data

Time-series graphs can not only communicate patterns in data about the past, but can also be used to help make inferences about the future. This chapter therefore considers how people's expectations about the future from time-series graphs might be influenced by characteristics of the plotted data, and by graph design/layout choices.

Study 7 explores how global trend direction (i.e. upward/positive and downward/negative slopes) and the direction of the most recent data in the time-series influence people's expectations about the future, finding that expectations exaggerate existing patterns. This study also identifies that individuals' expectations of the future show strong anchoring to linear trend lines fitted to data, eliminating effects of global trend direction and recent data. Further, the presence of a trend line narrows people's expectations towards the trend line.

Study 8 then investigates whether a more subtle visual cue of a directional arrow, (aligned with the slope of the line of best fit and placed over the start of the plotted data), produces similar effects as a trend line. The effects of global trend direction and the direction of the most recent data found in Study 7 are replicated. However, converse to the effect of a trend line, arrows were ineffectual in moderating expectations. The pattern of results across Studies 7 and 8 suggest that bottom-up perceptual saliency of visual features and/or top-down knowledge relating to the meaning of graph features may influence to what extent such features are drawn on when making inferences.

Study 9 investigates to what extent people's expectations about the future might be grounded in the spatial orientation of the plotted data by comparing horizontal graphs (with time running left-right) to vertical graphs (with time running top-bottom, aligned with the direction of gravity). Expectations about the future for vertical graphs follow the same patterns found in horizontal graphs indicating that interpretation of spatial relationships encoded in graphs are highly flexible in nature.

How do people form expectations about the future from time-series graphs?

Data represented in time-series graphs are often used to support decision-making. An investor might look at past share prices to try to infer future performance and decide whether to buy or sell. An epidemiologist might look at infection rates over time to help inform the rate at which the disease may spread into the future. Policy-makers might look at past global average temperatures to inform how temperatures may change into the future.

Indeed, a particularly contentious societal issue has been the apparent slowdown in global average surface temperatures since the late 1990s, commonly referred to as the global warming ‘pause’ or ‘hiatus’ (Hawkins, Edwards, & McNeall, 2014) (Figure 20). The debate has focused on whether the slowdown simply reflects short-term variability (i.e. noise) or a fundamental shift in the long-term trend (i.e. a change in signal) (Boykoff, 2014). The scientific community expected global average temperatures to continue to increase, attributing the ‘pause’ to short-term variability (Kerr, 2009; Lean & Rind, 2009).³ However, how did non-experts interpret the pattern in the data? When looking at graphs of the temperature data, both individuals who were sceptical of anthropogenic global warming (AGW) *and* individuals who accepted AGW made forecasts for subsequent years that were below the temperature rise than that indicated by the long-term trend (Lewandowsky, 2011). Neither group believed temperatures would stop rising, but nor did they believe that temperatures would continue to rise at the same rate as they had prior to the slowdown. It seems that certain patterns present in the data influenced people’s expectations.

³ Current evidence indicates that the slowdown was in fact an artefact in data collection, as ocean temperatures had been underestimated (Hausfather, et al., 2017; Karl, et al., 2015).

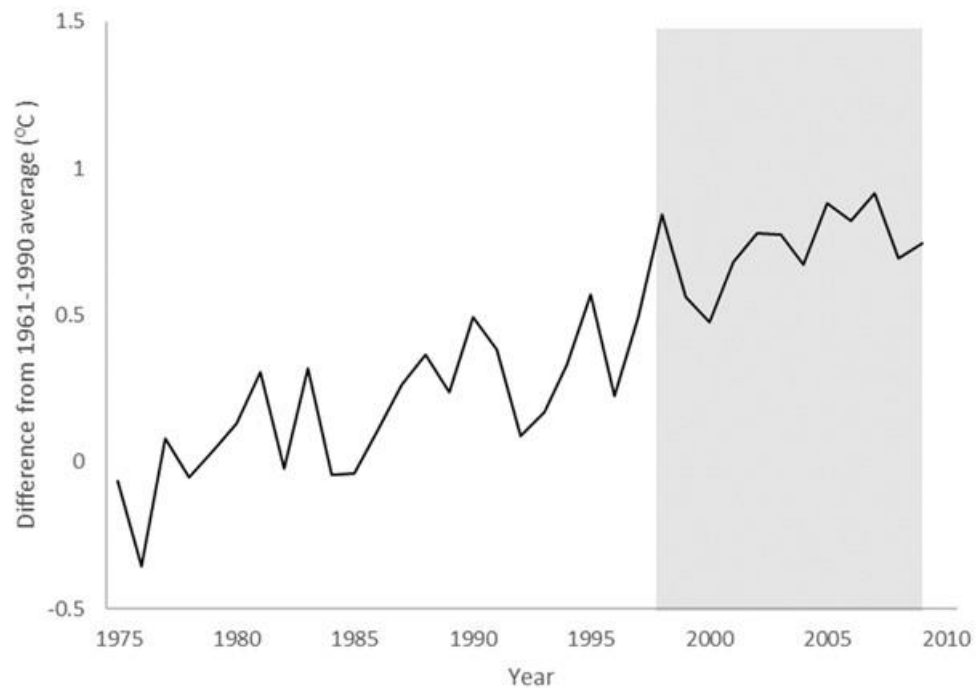


Figure 20. Yearly global mean surface temperature anomalies (relative to 1961–1990) between 1975 and 2009. Grey shaded area indicates period of supposed pause between 1998 and 2009. Data from HadCRUT4.

One such pattern might be the global long-term trend in a dataset. People tend to underestimate (Bolger & Harvey, 1993; Lawrence & Makridakis, 1989; Harvey & Reimers, 2013) or overestimate long-term trends (Harvey & Reimers, 2013), known as trend dampening and trend anti-dampening, respectively. The use of the anchoring heuristic (Tverksy & Kahneman, 1974) has been suggested to account for trend dampening (Bolger & Harvey, 1993). However, in a series of studies, forecasts showed trend anti-dampening for data showing negatively accelerating trends and linear trends with shallow slopes, whereas trend anti-dampening tended to be observed for data showing steeper trends (Harvey & Reimers, 2013). The authors of these studies propose that forecasts are not biased, but are adapted to patterns of trends that are typically experienced in the environment, in which time-series data typically shows cycles of growth and decay (Harvey & Reimers, 2013). This account is consistent with the growing literature suggesting that heuristic decision-making that deviates from

mathematically ‘rational’ choices are often effective choices in ecologically valid contexts (Gigerenzer & Gaissmaier, 2011).

Another pattern that might influence expectations about the future are short runs in data away from an existing long-term trend, such as the apparent global warming pause. Such local patterns at the end of datasets may be particularly salient, by virtue of being the most recent information. Salient information can be over-weighted when making judgements under uncertainty, distorting expectations (Schoemaker, 2004; Huber, Wider, & Huber, 1997). The most recent segment of time-series data (i.e. the connected line between the penultimate and last data point) has indeed been found to influence one-step ahead forecasting decisions (Lawrence & O’Connor, 1992; Bolger & Harvey, 1993).

More broadly, judgements made under uncertainty about runs in data have been shown to be dependent on the perceived ‘randomness’ of the data (Burns & Corpus, 2004; Ayton & Fischer, 2004). Runs attributed to random processes, such as three flips of a coin all landing on heads, tend to result in expectations for the next outcome to switch (i.e. tails) – known as the *gambler’s fallacy* (Kahneman & Tversky, 1972). Conversely, outcomes attributed to non-random processes, such as a basketball player successfully scoring three consecutive shots, tend to cause expectations for the next outcome to continue the run – known as the *hot hand* phenomenon (Gilovich, Vallone, & Tversky, 1985).

In time-series graphs, a short run of recent data that appears to depart from an existing trend might be interpreted as a meaningful (non-random) signal. Here expectations would be predicted to continue in the direction of the run, consistent with the hot hand phenomenon. Alternatively, the run might be attributed with short-term (random) variability. In this case, expectations would be predicted to switch direction, as per the gambler’s fallacy. For graphs of business-related data that show a greater random variability (noise), one-step ahead forecasts appear to show that the direction of the most recent segment results in forecasts consistent with the gambler’s fallacy, whereas when the data show less noise, forecasts were consistent with the hot-hand heuristic (Lawrence & O’Connor, 1992). However, it

is not known to what extent short runs of data away from underlying trends influence general expectations about the pattern of data into the future. An additional, as yet unexplored possibility is that a short run of data away from an underlying trend may increase uncertainty of how the data might evolve into the future, particularly when people are uncertain as to whether the run is caused by a random or non-random process.

The three studies reported in this chapter investigated three sets of questions. First, studies 7-9 investigate how global patterns (trends) and local patterns (recent data) influence people's expectations about the future. Studies 7 and 8 then ask whether the addition of visual features that provide information about the long-term trend (trend lines and arrows), influence these expectations. Study 9 then considers whether the spatial orientation of the plotted data, either horizontal with time running left-right or vertical with time running top-bottom, influences expectations.

In contrast to forecasting studies in which people are asked to make specific predictions, typically for the next time-point in a data-series or a set of sequential predictions into the future (Lawrence & O'Connor, 1992; Bolger and Harvey, 1993; Harvey & Reimers, 2013), the present studies used a novel paradigm to test expectations more generally. Participants were shown a possible future data point for the 25th time-point into the future (t_{25}) and were asked to make a judgement of whether they believed that data point was consistent or not with the past data. By probing different values for t_{25} across trials, a distribution of expectations was obtained. Hence, in contrast to making a forecast, the paradigm enables uncertainty over a range of possibilities to be captured.

Given that inferences made from visual displays are influenced by both bottom-up perceptual processing and top-down prior knowledge (Hegarty, 2011), the current set of studies limit potential effects of prior knowledge about the domain from which the data are drawn from by presenting fictional data on an obscure topic – namely, the luminosity (brightness) of stars in the galaxy plotted over time.

Study 7: Expectations with and without a trend line

This study was designed to investigate the extent of trend dampening and/or trend anti-dampening, and the influence of runs in the most recent data, on expectations of possible futures. Further, the influence of a linear trend line fitted to the data (using the least squares method) was also explored to see to what extent this might be used to anchor expectations about the future.

Features added to visual displays, such as trend lines, can be used to indicate particular characteristics of data. In the case of a linear trend line, the path of best fit through individual data points indicates to what extent the plotted data shows an increasing or decreasing trend, by virtue of the angle (slope) of the line (Bretscher, 2013). Instead of a reader having to mentally infer this trend, they can simply offload cognition onto perception of the plotted trend line. However, the addition of features that summarise statistical aspects of data can influence decision-making in unintended ways (Spiegelhalter, Pearson, & Short, 2011). For example, in visual displays of hurricane forecasts, people tend to focus on the most likely path of the hurricane (plotted as a solid line) and ignore the uncertainty around that path (plotted as an envelope encapsulating all ensemble forecast members), influencing decisions of whether to evacuate away from the hurricane or not (Broad, et al., 2007).

For time-series graphs with trend lines fitted, expectations about the future might be anchored on the trend lines rather than the underlying data. If so, a trend line might negate any trend dampening or anti-dampening effects. Further a trend line might influence how short runs of data away from the trend are interpreted. Short-runs of recent data may result in expectations to follow the hot-hand principle (i.e. to continue in the same direction) if the run is attributed to a non-random cause. However, adding a trend line may result in expectations to follow the gambler's fallacy (i.e. to switch direction). Given that a trend line will, on average, segment the plotted data equally above and below the trend, prior runs of data away from the trend line will return towards the trend. This may provide a salient cue indicating that such runs are the result of random variability.

Method

Design

To test expectations for time-series data, a forced choice task was employed in which participants were asked if a future data-point shown on a time-series graph was consistent or not with the plotted time-series. Confidence ratings for judgements were also collected. Graphs were presented as showing the luminosity of stars over time to limit effects of prior knowledge. The global long-term trend of the plotted data was either positive (i.e. an upward slope) or negative (i.e. a downward slope). The three data points corresponding to the most recent data in the plotted time-series, showed a local trend that matched the long-term trend (recent-consistent), showed a positive local trend (recent-up), or showed a negative local trend (recent-down). Graphs either showed a trend line through the plotted data, calculated using linear least squares regression (trend line), or did not show a trend line (no trend line), see Figure 21 for examples. The experiment was therefore a 2 (Global Trend Direction) x 3 (Recent Data) x 2 (Trend Line) design, with all variables within-participants.

Participants

Forty-one undergraduate students (32 female, 9 male) with normal or corrected to normal vision, from the University of East Anglia took part in the study in return for course credit or a nominal payment. Average age was 20 years (range 18-30 years). One participant was removed before data analysis as they withdrew from the study part way through. Sample size was informed via power analysis to detect a medium effect size ($\eta p^2 = .060$).

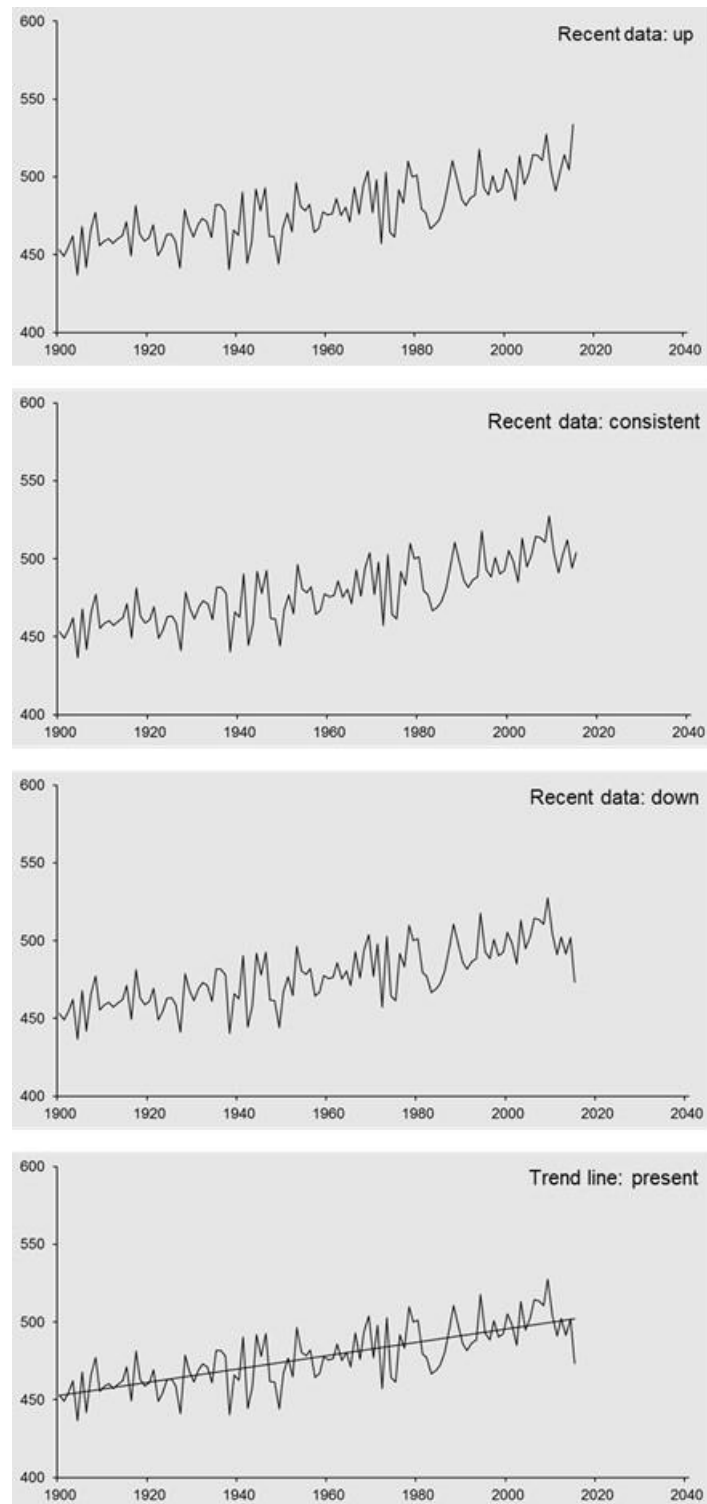


Figure 21. Examples of graph stimuli, showing the three levels of recent data for a graph with a positive global trend, and an example of trend line added.

Apparatus

TFT LCD monitors (51cm x 29cm) set to 1280 x 720 pixels were used for stimulus presentation and Eprime Version 2.0 (Psychology Software Tools Inc., Sharpsburg, USA) was used to control stimulus presentation and record data. Responses for judgements (yes/no) about the graphs were mapped to the ‘Z’ and ‘M’ keyboard keys, which were reversed and counterbalanced across participants. A visual analogue scale (VAS), controlled using the mouse was used to collect confidence ratings.

Graph Stimuli

Eighteen time-series graphs were created (1167 x 581 pixels), each plotting 116 data points (representing data for the years 1900-2015). The x-axis covered the years 1900 to 2040, meaning that no data was plotted for the years 2016-2040. The y-axis was labelled as “Luminosity (10^{24} Watts)” with units ranging 400-600.

Three initial time-series datasets were created. For each initial dataset, six variants were produced, reflecting the three different levels of recent data (recent-consistent, recent-up, recent-down) combined with the two levels of global long-term trends (positive, negative).

The initial time-series datasets plotted in the graphs contained upward (positive) global trends and consisted of an intercept, a global trend component, a noise (variability) component, and a recent data component. The plotted data took the form $y = 0.5t + 450 + \text{noise}$, where t is the time-point of the series. The noise component was created by sampling residuals at random from a normal distribution of $N(0,16)$, with the added criteria that no residual could be larger than ± 32 . Noise was added to the first 112 time-points. The residual for time-point 113 was set to zero.

Three variations of each initial dataset were created with different sets for the most recent data (time-points 114-116). In the recent-consistent variant, the residual for time-point 114 was sampled from the same normal distribution as before, but with the added criterion that it should be between 8-16 units; the residual for time-point 115 was set to the same as for time-point 114, but of the opposite sign. The residual for time-point 116 was set to zero. In the recent-up variant, residuals for time-points 114-116 were sampled from the same normal

distribution, but with the requirement that the residuals were between 8-16, 0-8 and 16-32, respectively. In the recent-down variant, residuals for time-points 114-116 were of the opposite sign of the recent-up residuals.

Datasets for downward (negative) global trends were created by mirroring each of the positive global trend datasets in the horizontal plane and setting the intercept to 550. Each cell of the design therefore contained graphs for three different datasets. Additional stimuli were created for practice trials in the same manner as described above.

Trend line

Graphs in the trend line present condition were identical to the trend line absent condition with the exception that a straight line was fitted to the data using linear least square regression. Across graphs, the slope of the trend line varied between ± 0.41 - 0.51 (range due to the random sampling of the noise component). Trend lines were plotted with the same line weight as the plotted data.

Graph probes

For each graph, seven probe data-points were determined for the 140th time-point in the series, (year 2040, i.e. 25 time-points into the future, t_{25}). A central probe simply fell on the global long-term trend. The remaining six probes were set either side of the central probe, representing the upper and lower 95%, 99.5% and 99.99% limits of the noise component (corresponding to ± 1.98 , ± 2.86 and ± 4.03 SDs of the distribution of the noise). Probes were plotted as a cross (Figure 22).

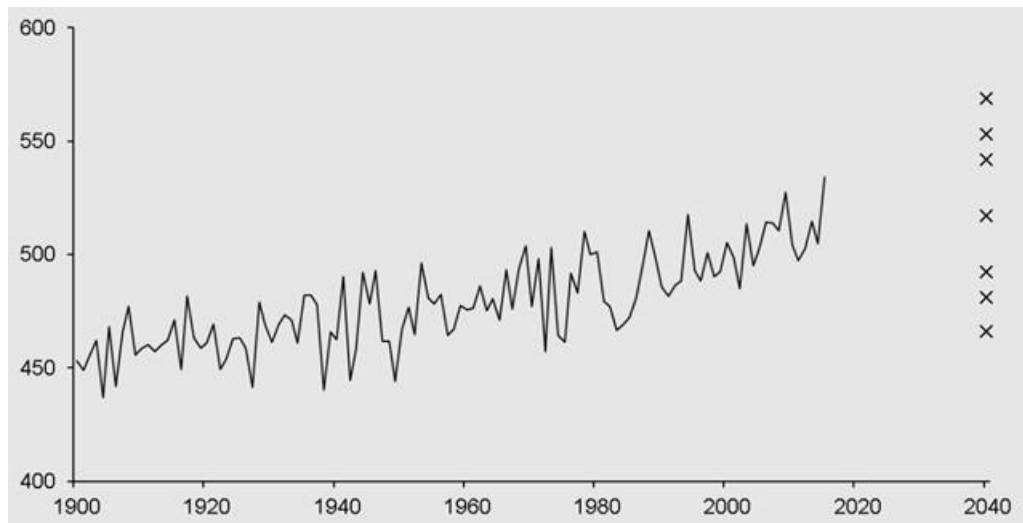


Figure 22. Example of the probe locations in relation to one of the graph stimuli.
Note: only one probe appeared on any given trial in the experiment.

Procedure

Participants were informed that the study was investigating how people interpret line graphs and they then received instructions on screen before completing a practice block of trials. Each trial consisted of a study phase during which participants were asked to look at a graph. This was followed by a judgement phase in which one of the probes appeared on the graph indicating a possible estimate of the luminosity of the star in the year 2040 (Figure 23). Participants' task was to indicate a 'yes – consistent' or 'no – not consistent' response as to whether the estimate for 2040 was consistent or not with the plotted data for 1900-2015. Participants were instructed to give a response as quickly and accurately as possible when the probe appeared. If no response was given within 2 seconds, a prompt appeared on screen asking for a response, and a reminder was shown encouraging a faster response on future trials. Participants were then asked to rate how confident they were in their judgement using a visual analogue scale (VAS), labelled from 'not confident at all' through to 'extremely confident'.

Trials were presented in two blocks, one block consisting of all the trend line absent trials and the other block consisting of the trend line present trials.

Each block contained 126 trials (3 datasets x 3 levels of recent data x 2 levels of global long-term trends x 7 probe locations). Block order was counterbalanced across participants. The order of trials within each block was randomised, but with the condition there were no consecutive trials with the same global long-term trend. The study lasted approximately 50 minutes.

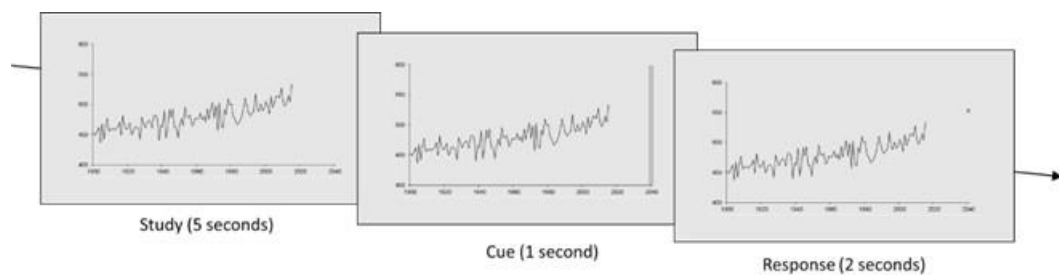


Figure 23. Summary of a trial.

Results

Individual trials were removed from analysis if reaction times from the onset of the probe were less than 200ms or ± 3 standard deviations from the participant's mean reaction time (1.54% of trials removed).

Changes in mean location of expected future values

To determine whether the independent variables influenced expectations of the future trajectory of time-series, the mean location of expectations at t_{25} was calculated for each participant and for each cell of the experimental design. This value was determined using the distribution of responses across the seven probe locations, sampled across the three different datasets for each cell of the study design. For each cell of the study design and for each participant, the 'yes' response rate at each probe location was first calculated – this was simply the proportion of trials that were responded with a 'yes' response from the total number of trials where a response was given (Figure 24). Then the 'yes' response

rate at each probe location was multiplied by the value of the probe location (in units of SDs of the distribution of the noise). This weights each probe location by the ‘yes’ response rate. The sum of the response rate, and the sum of the weighted probe locations were then calculated. The mean location was then calculated as the sum of the weighted probe locations divided by the sum of the ‘yes’ response rates. Figure 24 shows a worked example of these calculations.

Plotted probes	Probe location	‘Yes’ responses	‘Yes’ response rate	‘Yes’ response rate multiplied by probe location (weighted probe location)
×	+4.03	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/>	2 / 3 = 0.67	+4.03 x 0.67 = 2.69
×	+2.86	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	3 / 3 = 1.00	+2.86 x 1.00 = 2.86
×	+1.98	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	3 / 3 = 1.00	+1.98 x 1.00 = 1.98
×	0.00	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/>	2 / 3 = 0.67	0.00 x 0.67 = 0.00
×	-1.98	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	1 / 3 = 0.33	-1.98 x 0.33 = -0.65
×	-2.86	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	0 / 3 = 0.00	-2.86 x 0.00 = 0.00
×	-4.03	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	0 / 3 = 0.00	-4.03 x 0.00 = 0.00
			$\Sigma = 3.67$	$\Sigma = 6.88$

2020 2040

$$\text{Mean location: } 6.88 / 3.67 = 1.87$$

Figure 24. Worked example of the calculation of the mean location of expectations for one participant and one cell of the study design. Each box in the ‘Yes’ responses column represents a single trial where a response was given; crosses in boxes indicates a ‘yes’ response to that trial; empty boxes indicates a ‘no’ response to that trial.

Mean locations of expectations across conditions were then compared by submitting scores to a 2 (Trend Line [absent, present]) x 2 (Global Trend Direction [positive, negative]) x 3 (Recent Data [recent-consistent, recent-up, recent-down]) x 2 (Block Order [trend line absent-present, trend line present-absent]) mixed ANOVA (Table 12). Means and standard deviations for each cell of the analysis are provided in Appendix 6, Table A6-1.

There was a main effect of global trend direction, in which there was a positive bias for positive trends ($M = 0.310$, $SD = 0.649$) and a negative bias for negative trends ($M = -0.393$, $SD = 0.591$); $d = 1.133$, 95% CI [0.549, 1.706]. There was also a main effect of recent data, in which there was a positive bias in the recent-up condition ($M = 0.385$, $SD = 0.447$) and a negative bias for the recent-down condition ($M = -0.468$, $SD = 0.401$), which were both different to the recent-consistent condition ($M = -0.041$, $SD = 0.393$); $t(39) = 7.772$, $p < .001$, $d = 1.010$, [0.668, 1.346], $t(39) = -7.157$, $p < .001$, $d = -1.076$, [-1.450, -0.693], respectively.

Table 12. Study 7 mixed ANOVA for changes in mean location of expected future values; * indicates significance at the .05 level. † indicates Greenhouse-Geisser corrected statistic.

Source	Test	<i>p</i>	ηp^2
Main effects			
trend line	$F(1,38) = 3.476$.070	.084
global trend direction	$F(1,38) = 17.784$	< .001*	.319
recent data	$F(2,76) = 86.934^\dagger$	< .001*	.696
block order	$F(1,38) = 0.371$.546	.010
Two-way interactions			
trend line x global trend direction	$F(1,38) = 32.493$	< .001*	.461
trend line x recent data	$F(2,76) = 52.610$	< .001*	.581
trend line x block order	$F(1,38) = 1.840$.183	.046
global trend direction x recent data	$F(2,76) = 0.110$.896	.003
global trend direction x block order	$F(1,38) = 0.228$.635	.006
recent data x block order	$F(2,76) = 3.343$.041*	.081
Three-way interactions			
trend line x global trend direction x recent data	$F(2,76) = 0.225$.799	.006
trend line x global trend direction x block order	$F(1,38) = 1.222$.276	.031
trend line x recent data x block order	$F(2,76) = 9.937$	< .001*	.207
global trend direction x recent data x block order	$F(2,76) = 2.720$.072	.067
Four-way interaction			
trend line x global trend direction x recent data x block order	$F(2,76) = 0.538$.586	.014

There was a significant two-way interaction between trend line and global trend direction in which trend lines reduced trend anti-dampening (Figure 25). Post-hoc tests revealed a significant difference between positive and negative global trend directions when there was no trend line, $t(39) = 5.225$, $p < .001$, $d = 1.377$, 95% CI [0.771, 1.970]; ($M = 0.416$, $SD = 0.827$, and $M = -0.636$, $SD =$

0.695, respectively). However the size of this effect was reduced when there was a trend line, $t(39) = 2.418$, $p = 0.020$, $d = 0.581$, [0.089, 1.066]; ($M = 0.205$, $SD = 0.612$, and $M = -0.150$, $SD = 0.609$, respectively); note: t -test is not significant with a corrected α for multiple comparisons ($\alpha = .05/4 = .0125$). Further, there was a significant difference between trend line absent and trend line present conditions when the global trend direction was negative $t(39) = 5.503$, $p < .001$, $d = 0.743$ [0.428, 1.052]; and a similar, but smaller effect size when the global trend direction was positive $t(39) = 2.037$, $p = .048$, $d = 0.291$ [0.576, 0.002]; note t -test is not significant with a corrected α for multiple comparisons ($\alpha = .05/4 = .0125$).

There was also a significant two-way interaction between trend line and recent data in which trend lines mitigated continuation of runs (Figure 26). Post-hoc tests found a significant difference between trend line absent and trend line present trials with recent-down data $t(39) = 7.013$, $p < .001$, $d = 1.305$, 95% CI [0.835, 1.765]; ($M = -0.791$, $SD = 0.501$, and $M = -0.144$, $SD = 0.490$, respectively) and with recent-up data, $t(39) = 3.731$, $p = .001$, $d = -0.746$ [-1.168, -0.317] ($M = 0.600$, $SD = 0.644$, and $M = 0.169$, $SD = 0.500$, respectively), but there was no reliable difference with recent-consistent data, $t(39) = 2.123$, $p = .040$, $d = 0.399$ [0.018, 0.775] ($M = -0.138$, $SD = 0.496$, and $M = 0.057$, $SD = 0.482$); note t -test is not significant with a corrected α for multiple comparisons ($\alpha = .05/4 = .0125$).

This two-way interaction was further moderated by a three-way interaction involving block order. The pattern of data was the same for both block orders as per the two-way interaction between trend line and recent data described above, except that with recent-up data there was no significant difference between trend line absent trials and trend line present trials when participants received trend line present trials prior to trend line absent trials ($M = 0.286$, $SD = 0.622$; $M = 0.225$, $SD = 0.418$, respectively; $t(19) = 0.428$, $p = 0.674$, $d = -0.155$ [-0.642, 0.415]), indicating the possibility of transfer effect of the influence of a trend line.

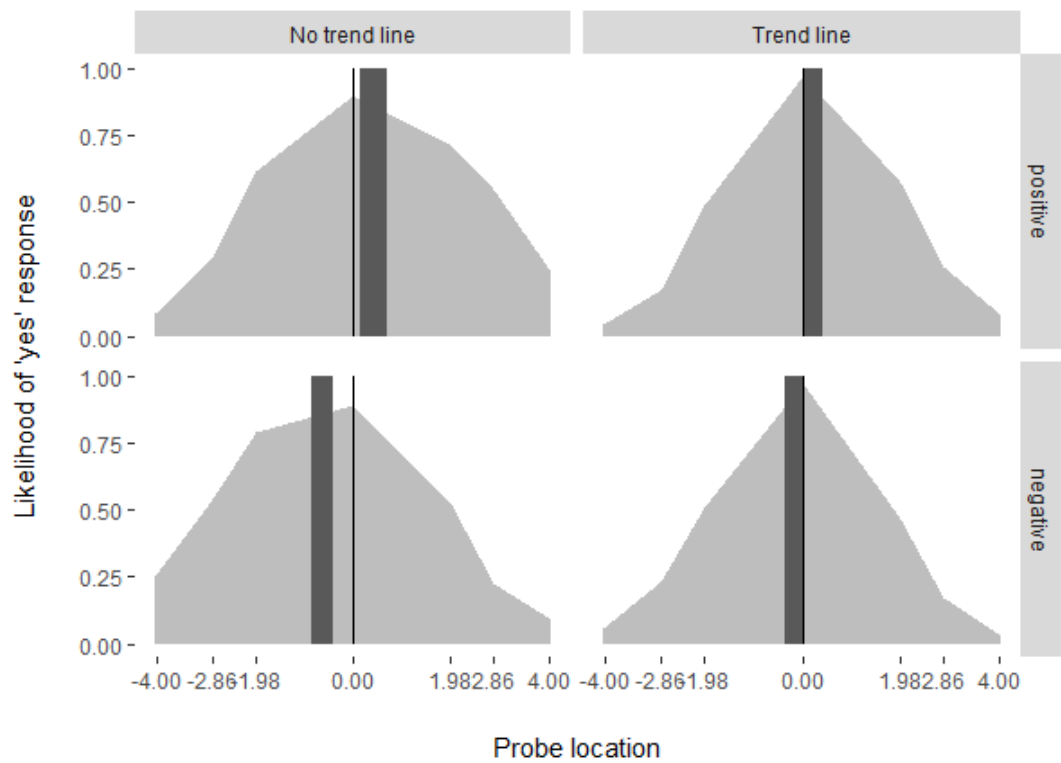


Figure 25. Interaction between trend line and global trend direction. Vertical dark grey bars indicate the 95% confidence interval for the mean location of expectation distributions. Vertical line at probe location 0 provided as a reference point. Light grey shaded areas indicate the full distribution of 'yes' responses for each condition.

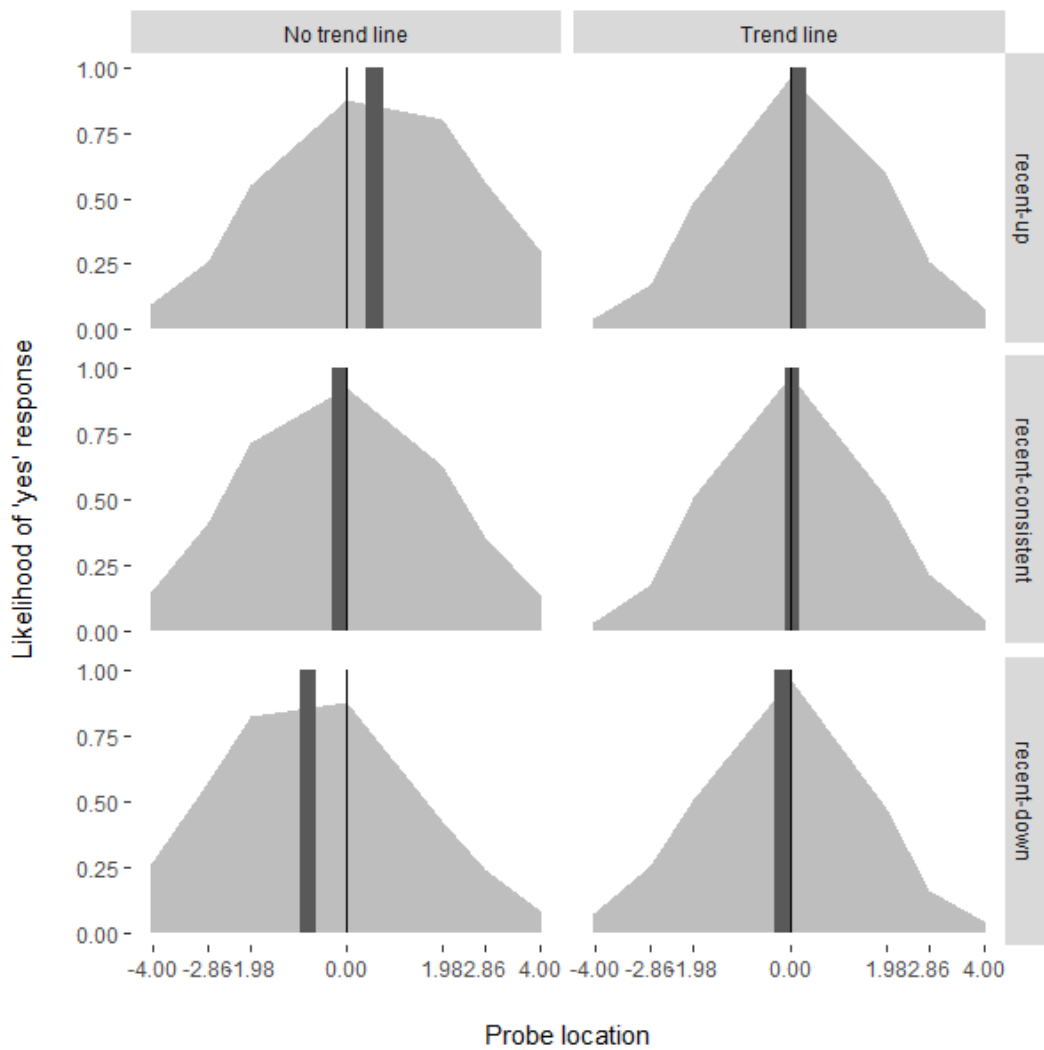


Figure 26. Interaction between trend line and recent data. Vertical dark grey bars indicate the 95% confidence interval for the mean location of expectation distributions. Vertical line at probe location 0 provided as a reference point. Light grey shaded areas indicate the full distribution of 'yes' responses for each condition.

Expectations for future values outside 95% range

The results above suggest that the presence of a trend line reduces trend anti-dampening and reduced continuation of recent runs in data. This effect of the trend line could be to narrow expectations – i.e. rejecting probes above and below the trend line in equal measure, or to widen expectations – i.e. accepting probes above and below the trend line in equal measure. As shown in Figure 25 and Figure 26 above, the data clearly suggest that responses were narrowed in the presence of a trend line. As a check, the mean of the number of ‘yes’ responses to the six outer probes (representing the tails outside the 95% spread of the plotted data) were submitted to a 2 (Trend Line [absent, present]) x 2 (Block Order [trend line absent-present, trend line present-absent]) mixed ANOVA (Table 13). Means and standard deviations for each cell of the analysis are provided in Appendix 6, Table A6-2.

Table 13. Study 7 mixed ANOVA for mean acceptance rates of the outer probes; * indicates significance at the .05 level.

Source	Test	<i>p</i>	ηp^2
Main effects			
trend line	$F(1,38) = 53.895$	< .001*	.568
block order	$F(1,38) = 1.430$.239	.036
Two-way interaction			
trend line x block order	$F(1,38) = 6.615$.014*	.148

There was a main effect of trend line, where there was a greater response rate to outer probes when the trend line was absent ($M = 0.411$, $SD = 0.117$) than when the trend line was present ($M = 0.256$, $SD = 0.134$); $t(39) = 6.864$, $p < .001$, $d = 1.229$, 95% CI [0.780, 1.667]. There was no main effect of block order, but there was a two-way interaction between block order and trend line. Post-hoc tests found significant differences in outer probe response rates between the trend line absent and trend line present conditions for both block orders (absent-present,

$t(19) = 6.072, p < .001, d = 1.784, [0.947, 2.525]$; present-absent, $t(19) = 4.129, p = .001, d = 0.797 [0.335, 1.244]$. There was no difference in response rates to outer probes between the two block orders when the trend line was absent, $t(38) = 0.406, p = 0.687$, and no reliable difference when the trend line was present $t(38) = -2.313, p = .026$ (note: corrected α for multiple comparisons, $\alpha = .05/4 = .0125$).

Confidence in judgements

To understand whether the inclusion of a trend line influences confidence in judgments, mean VAS scores for trials in which the probe was judged consistent with past data were submitted to a 2 (Trend line [absent, present]) x 2 (Probe location [central, outer])⁴ x 2 (Block Order [trend line absent-present, trend line present-absent]) mixed ANOVA.⁵ Lower VAS scores map to lower confidence and higher VAS scores map to higher confidence (min = 0, max = 100). Means and standard deviations for each cell of the analysis are provided in Appendix 6, Table A6-3.

There were main effects of trend line and probe location, which were further moderated by significant two-way interactions (Table 14). The interaction between trend line and block order revealed that when graphs without trend lines were presented first, confidence scores increased when graphs with trend lines were subsequently presented, $t(18) = 3.763, p = .001, d = 0.557, 95\% \text{ CI } [0.210, 0.893]$ ($M = 65.299, SD = 11.515; M = 72.183, SD = 13.150$, respectively). Conversely there was no difference between conditions when graphs with trend lines were presented first ($M = 71.252, SD = 14.542; M = 70.151, SD = 11.325$, respectively; $t(19) = -0.813, p = .426, d = -0.085, [-0.289, 0.122]$). Results therefore suggest a transfer effect, such that making judgements about graphs with trend lines leads to greater confidence on subsequent judgements about graphs that don't have trend lines.

⁴ 'Central' relates to the middle probe, 'outer' relates to the six probes either side of the middle probe.

⁵ Data were collapsed across global trend and recent data.

Critically, the interaction between trend line and probe location revealed that when the trend line was present, there was greater confidence in ‘yes’ judgements for central probes compared to when the trend line was absent, $t(39) = 7.197$, $p < .001$, $d = 0.994$, [0.642, 1.339]. Conversely, for outer probes there was less confidence in ‘yes’ responses when the trend line was present compared to when it was absent, $t(38) = -3.809$, $p < .001$, $d = -0.438$, [-0.681, -0.190].

Table 14. Study 7 mixed ANOVA for mean confidence ratings on ‘yes’ judgements; * indicates significance at the .05 level.

Source	Test	p	ηp^2
Main effects			
trend line	$F(1,37) = 6.543$.015*	.150
probe location	$F(1,38) = 181.866$	<.001*	.831
block order	$F(1,37) = 0.251$.619	.007
Two-way interactions			
trend line x probe location	$F(1,37) = 60.352$	<.001*	.620
trend line x block order	$F(1,37) = 12.478$.001*	.252
probe location x block order	$F(1,37) = 0.035$.853	.001
Three-way interaction			
trend line x probe location x block order	$F(1,37) = 0.720$.401	.019

Discussion

In the absence of a trend line, expectations for data with positive global trends were weighted above the long-term trend, while expectations for data with negative global trends were weighted below the long-term trend. Hence, expectations showed trend anti-dampening, consistent with prior research using graphs with shallow trends (Harvey & Reimers, 2013). However, when a trend line was added, no trend anti-dampening was observed, expectations narrowed (i.e. the range of values considered consistent with past data was smaller) and people had greater confidence in expectations that fell in line with the trend line

and reduced confidence in expectations that fell either side of the trend line.

People's expectations therefore appear to be anchored to trend lines.

People's expectations were also sensitive to short-runs of recent data away from global trends. In the absence of a trend line, expectations shifted in the direction of the run, and away from the global trend in the data. This suggests that people may anchor expectations on recent data and attribute such runs as changes in signal, rather than attributing them solely to random short-term variability (noise), consistent with the hot-hand phenomenon (Gilovich, Vallone, & Tversky, 1985) and with one-step ahead forecasts made from time-series data (Lawrence & O'Connor, 1992). Weighting expectations on the most recent data might be an adaptive response, either to time-series data (e.g. if time-series generally tend to show positive serial correlation (see Jebb, Tay, Wang, & Huang, 2015) or more generally to runs (Tyszka, et al., 2017).

However, the presence of a trend line significantly mitigated the effects of runs of recent data, causing expectations to be more in line with global trends. Here, the run of recent data may be anchored on to inform expectations, but with the trend line acting as a salient frame of reference to indicate that runs of data are random in nature and so may not continue. Conversely, expectations may simply be anchored on the trend line with only minor adjustments made in response to runs.

Study 8: Expectations with and without an informational arrow

Finding that expectations are sensitive to the global trend direction and recent runs of data, and that trend lines mitigate these sensitivities, the present study next investigates whether cueing the slope of the global trend, by way of a directional arrow, informs expectations about the future in a similar way to directly representing the trend line. Explicitly showing the trend line in a figure could be regarded as attempting to lead the viewer to a particular conclusion, whereas an arrow merely indicates that the trend might be relevant. More broadly, it is desirable to reduce clutter on graphs (Rosenholtz, Li, & Nakano, 2007; Tufte, 2006), while also supporting relevant inferences. If a trend line provides a frame of reference against which the plotted data is interpreted, then cueing cognition for the slope of the global trend using an arrow may result in similar effects on expectations.

Conversely, if expectations are simply anchored on trend lines, by virtue of their perceptual saliency, then cueing cognition for the slope of the trend may not affect expectations. To explore these possibilities, Study 7 was replicated, but instead of providing a trend line, an arrow was presented at the start of the time-series, the angle of which was aligned with the slope of the global trend.

Interpreting a static image to make inferences about the future states is an example of perceptual simulation (Coventry, et al., 2010; Coventry, et al., 2013; Hegarty & Simms, 1994; Hegarty, 1992), and more generally, mental simulation – i.e. forming predictions about future events to plan and adapt behaviour in anticipation (Atance & O’Neill, 2001; Suddendorf & Corballis, 2007; Decety & Grèzes, 2006; Schacter, Addis, & Buckner, 2007). In static images, arrows can convey a meaning of a change over time, of a direction of movement, and of a path (Tversky, 2011), and may support perceptual simulation by augmenting cognition (Heiser & Tversky, 2006; Freyd & Pantzer, 1995). Furthermore, arrows are salient symbolic elements that can direct attention automatically (Hommel, et al., 2001; Tipples, 2002; Pratt & Hommel, 2003; Kuhn & Kingstone, 2009). Therefore, when aligned with the slope of global trends, arrows might be effective visual cues to support mental inferences about long-term trends in data. If so,

arrows should mitigate effects of global trends and runs in recent data on future expectations, similar to trend lines.

Method

Design

The same forced choice task as Study 7 was employed. Graphs either showed an arrow over the leftmost plotted data, the direction of which was aligned to the trend of the data using linear least squares regression (arrow present) (Figure 27), or did not show an arrow (arrow absent). As with Study 7, the direction of the global trend (positive or negative) and the direction of the recent-data (end-consistent, end-up, or end-down) was also manipulated. The experiment was therefore a 2 (Arrow) x 2 (Global Trend Direction) x 3 (Recent-Data) design, with all variables within participants.

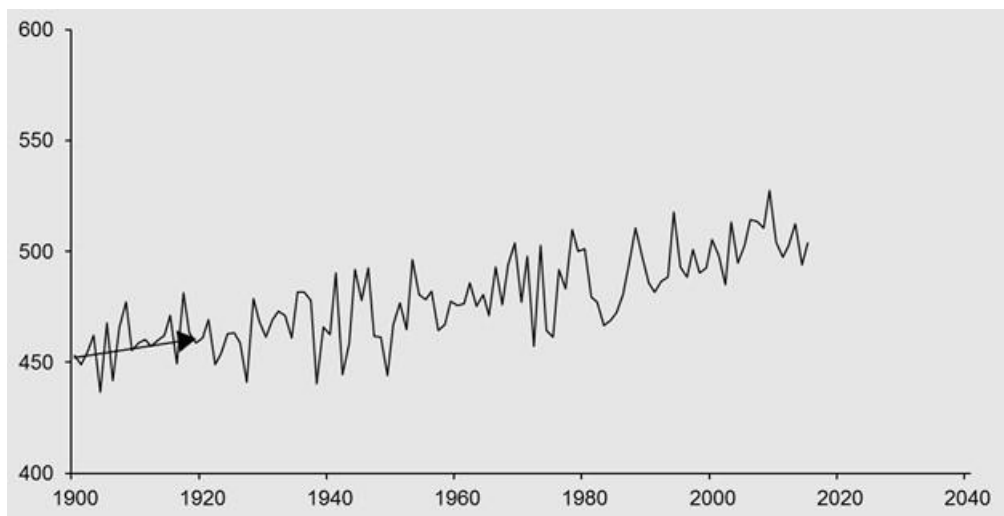


Figure 27. Example of ‘arrow present’ stimuli.

Participants

Thirty-two undergraduate students (21 female, 11 male) with normal or corrected to normal vision, from the University of East Anglia took part in the study in

return for course credit or a nominal payment. Average age was 21 years (range 18-36 years).

Apparatus

Apparatus and response key mappings were as per Study 7.

Graph Stimuli, Arrows and Probes

Graph stimuli for the arrow absent trials were the same as the trend line absent stimuli used in Study 7. A version of each of the graphs was created in which an arrow was placed at the start of the time-series data and angled and pointed in line with direction of the long-term trend. Graph probes were calculated as per Study 7.

Procedure

The procedure was the same as per Study 7. The study lasted approximately 50 minutes.

Results

Screening criteria was as per Study 7, resulting in 1.43% of trials removed from further analysis. Response distributions were calculated for each cell of the experimental design as per Study 7.

Changes in mean location of expected future values

The mean of the response distributions across conditions was compared by submitting mean scores to a 2 (Arrow [absent, present]) x 2 (Global Trend Direction [positive, negative]) x 3 (Recent Data [end-consistent, end-up, end-down]) x 2 (Block Order [arrow absent-present, arrow present-absent]) mixed ANOVA (Table 15). Means and standard deviations for each cell of the analysis are provided in Appendix 7, Table A7-1.

There was a main effect of global trend direction, where there was a positive bias for positive trends ($M = 0.283$, $SD = 0.675$) and a negative bias for negative trends ($M = -0.326$, $SD = 0.695$); $t(31) = -2.917$, $p = .007$, $d = -0.890$,

95% CI [-1.521, -0.246]. There was a main effect of recent data, where there was a greater positive bias for recent-up data ($M = 0.625$, $SD = 0.558$) and a greater negative bias for recent-down data ($M = -0.664$, $SD = 0.499$), when compared with recent-consistent data ($M = -0.025$, $SD = 0.368$); $t(31) = 7.409$, $p < .001$, $d = 1.375$, [0.871, 1.868], $t(31) = -9.066$, $p < .001$, $d = -1.457$ [-1.931, -0.973], respectively. However, there was no interaction between global trend direction and arrow (Figure 28) and no interaction between recent data and arrow (Figure 29).

Table 15. Study 8 mixed ANOVA for changes in mean location of expected future values; * indicates significance at the .05 level. † indicates Greenhouse-Geisser corrected statistic.

Source	Test	<i>p</i>	ηp^2
Main effects			
arrow	$F(1,30) = 0.105$.749	.003
global trend direction	$F(1,30) = 8.307$.007*	.217
recent data	$F(2,60) = 80.367^\dagger$	< .001*	.728
block order	$F(1,30) = 0.287$.596	.009
Two-way interactions			
arrow x global trend direction	$F(1,30) = 0.163$.690	.005
arrow x recent data	$F(2,60) = 2.432$.096	.075
arrow x block order	$F(1,30) = 2.323$.138	.072
global trend direction x recent data	$F(2,60) = 1.842$.167	.058
global trend direction x block order	$F(1,30) = 0.259$.614	.009
recent data x block order	$F(2,60) = 1.374$.261	.044
Three-way interactions			
arrow x global trend direction x recent data	$F(2,60) = 1.253$.293	.040
arrow x global trend direction x block order	$F(1,30) = 3.530$.070	.105
arrow x recent data x block order	$F(2,60) = 2.735$.073	.084
global trend direction x recent data x block order	$F(2,60) = 1.467$.239	.047
Four-way interaction			
arrow x global trend direction x recent data x block order	$F(2,60) = 0.763$.471	.025

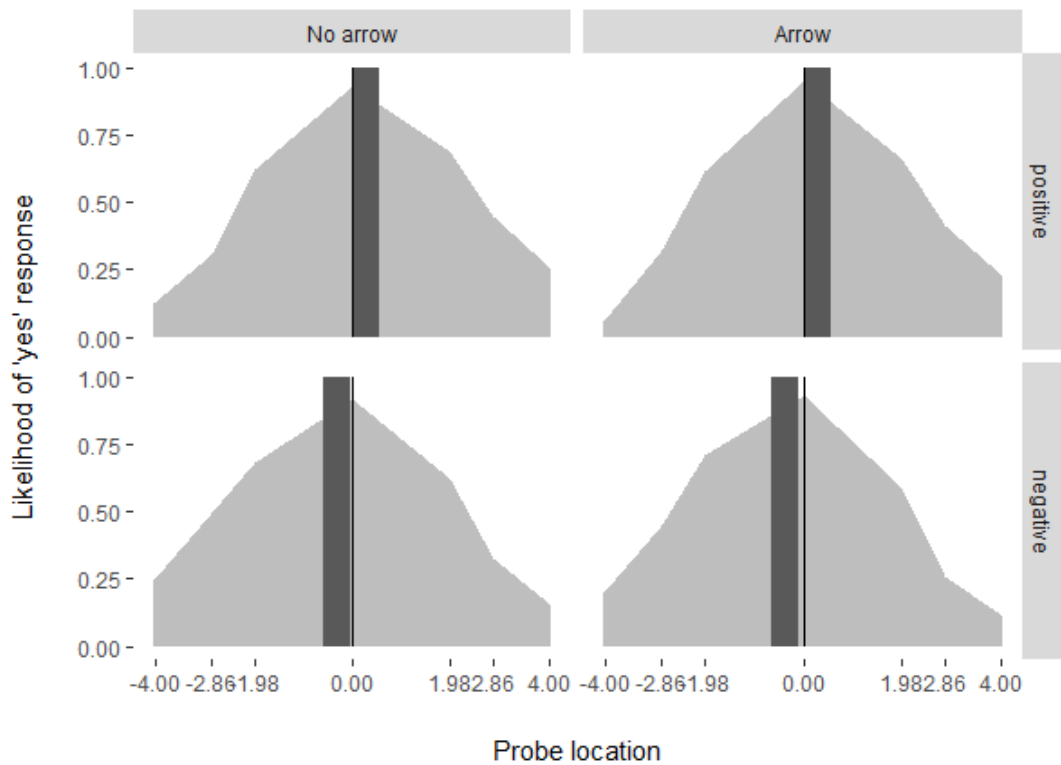


Figure 28. No interaction between arrow and global trend direction. Vertical dark grey bars indicate the 95% confidence interval for the mean location of expectation distributions. Vertical line at probe location 0 provided as a reference point. Light grey shaded areas indicate the full distribution of 'yes' responses for each condition.

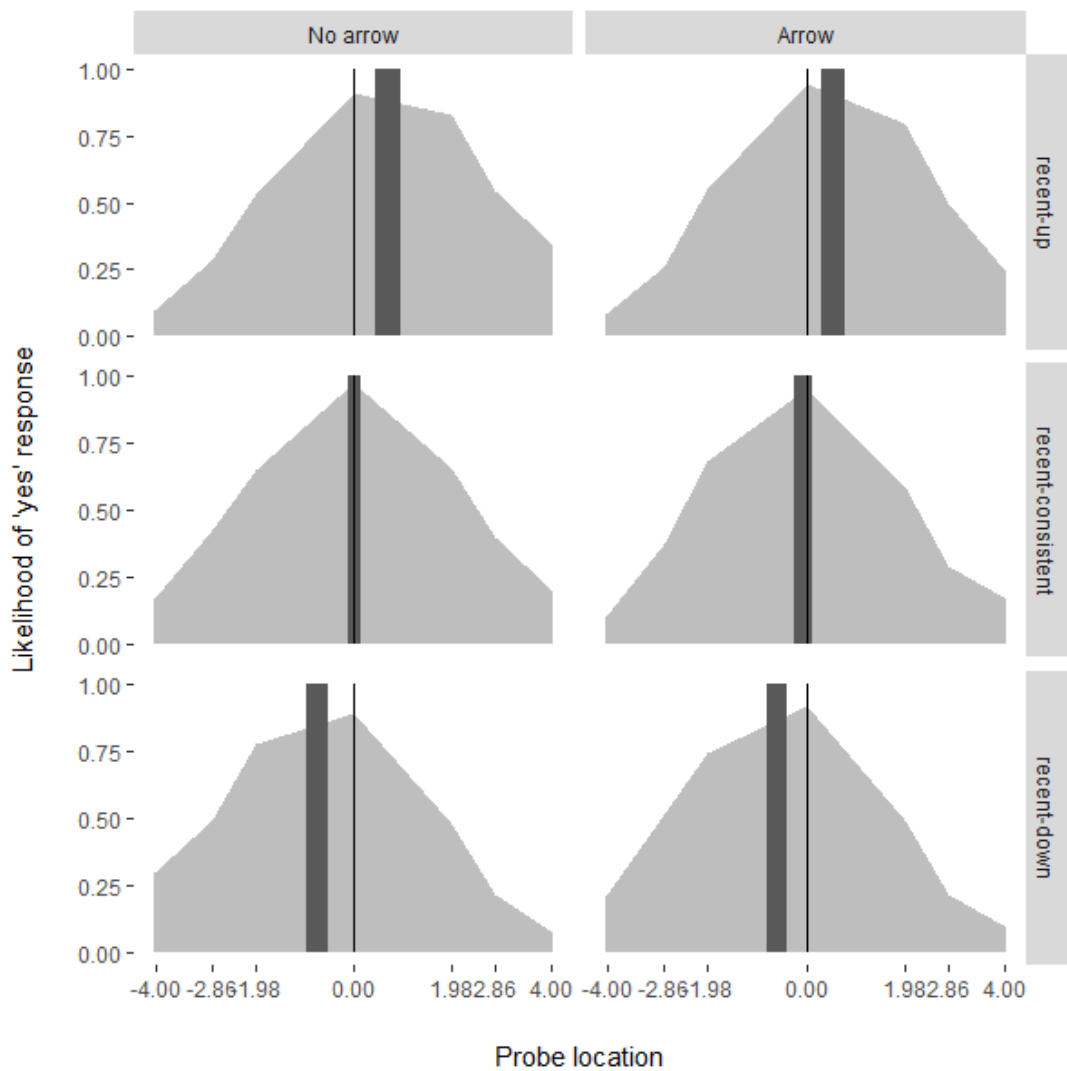


Figure 29. No interaction between arrow and recent data. Vertical dark grey bars indicate the 95% confidence interval for the mean location of expectation distributions. Vertical line at probe location 0 provided as a reference point. Light grey shaded areas indicate the full distribution of 'yes' responses for each condition.

Expectations for future values outside 95% range

Consistent with Study 7, mean ‘yes’ responses to the probes outside the 95% range were compared by submitting scores to a 2 (Arrow [absent, present]) x 2 (Block Order [arrow absent-present, arrow present-absent]) mixed ANOVA (Table 16). Means and standard deviations for each cell of the analysis are provided in Appendix 7, Table A7-2.

There was a main effect of arrow, where there was a greater response rate to outer probes when the arrow was absent ($M = 0.414$, $SD = 0.124$) than when the arrow was present ($M = 0.384$, $SD = 0.131$); $t(31) = 2.501$, $p = .018$, $d = 0.236$, 95% CI [0.040, 0.428]. There was no main effect of block order, but there was a two-way interaction between block order and arrow. Post-hoc tests found a significant difference between arrow absent and arrow present outer probe response rates in the absent-present block order, $t(15) = 3.131$, $p = .007$, $d = 0.477$, [0.128, 0.814]; but not in the present-absent block order, $t(15) = 0.267$, $p = .793$, $d = 0.026$, [-0.167, 0.219]. The data are consistent with a narrowing of expectations in response to an arrow only when individuals first saw no arrow trials, indicating that the effect of an arrow here may be a demand characteristic.

Table 16. Study 8 mixed ANOVA for mean acceptance rates of the outer probes; * indicates significance at the .05 level.

Source	Test	p	ηp^2
Main effects			
arrow	$F(1,30) = 7.180$	< .012*	.193
block order	$F(1,30) = 0.838$.367	.027
Two-way interaction			
arrow x block order	$F(1,30) = 5.586$.025*	.157

Confidence in judgements

As per Study 7, mean VAS scores for trials in which the probe was judged consistent with past data were submitted to a 2 (Arrow [absent, present]) x 2 (Probe location [central, outer]) x 2 (Block Order [trend line absent-present, trend

line present-absent]) mixed ANOVA (Table 17). Means and standard deviations for each cell of the analysis are provided in Appendix 7, Table A7-3.

There was a main effect of probe location, in which confidence scores were greater for central probes than outer probes ($M = 77.987$, $SD = 13.291$; $M = 64.791$, $SD = 12.272$, respectively; $t(31) = 7.087$, $p < .001$, $d = 1.032$, 95% CI [0.644, 1.410]). There was also a main effect of arrow, which was moderated by an interaction with block order. The interaction between arrow and block order revealed that when no arrow graphs were presented first, confidence scores increased when arrow graphs were subsequently presented ($M = 66.723$, $SD = 13.886$; $M = 75.818$, $SD = 11.417$, respectively; $t(15) = 4.978$, $p < .001$, $d = 0.716$, [0.330, 1.088]). Conversely, there was no difference between arrow graphs and no arrow graphs when arrow graphs were presented first ($M = 71.317$, $SD = 13.167$; $M = 71.698$, $SD = 10.720$, respectively; $t(15) = 0.225$, $p = .825$, $d = 0.032$, [-0.246, 0.309]). Consistent with the addition of a trend line in Study 7, results indicate a transfer effect in which making judgements about graphs containing an arrow leads to greater confidence on expectation judgements for graphs without arrows.

Table 17. Study 8 mixed ANOVA for mean confidence ratings on ‘yes’ judgements; * indicates significance at the .05 level.

Source	Test	p	ηp^2
Main effects			
arrow	$F(1,30) = 14.440$.001*	.325
probe location	$F(1,38) = 50.132$	<.001*	.626
block order	$F(1,37) = 0.003$.955	<.001
Two-way interactions			
arrow x probe location	$F(1,30) = 2.291$.141	.071
arrow x block order	$F(1,37) = 12.209$.002*	.289
probe location x block order	$F(1,37) = 0.942$.339	.030
Three-way interaction			
arrow x probe location x block order	$F(1,37) = 1.132$.296	.036

Discussion

In the absence of an arrow, expectations showed trend anti-dampening and continuation of runs of recent data, replicating the effects found in Study 7. However, in contrast to trend lines, there was no evidence to suggest that arrows mitigate trend anti-dampening or continuation of runs. Although there was some evidence to suggest that arrows narrowed the range of expectations about the future, this was only evident when the block of arrow trials were presented after the block of no arrow trials, indicating a possible demand characteristic. While arrows do not appear to influence the distribution of future expectations, there was some evidence to suggest that they might increase people's confidence in expectations.

The data across Study 7 and Study 8 suggest that trend lines, but not arrows, influence expectations. Why might this be? One possibility is that trend lines in time-series graphs have a generally well-defined meaning of indicating the slope of a global trend. Conversely, the meaning of the arrow may not have been intuitively understood. Arrows can convey multiple meanings, and so their meaning in the context of the graphs may be ambiguous (Tversky, 2011), thereby limiting the strength of their attentional effects (Friesen, Ristic, & Kingstone, 2004; Gibson & Bryant, 2005).

Another possibility is that if people's expectations anchor on the most recent information (Lawrence & O'Connor, 1992; Bolger and Harvey, 1993), trend lines, may be effective because they were plotted across the most recent information. As the arrows were placed at the start of the plotted data, they may not have captured visual attention, or their informational content may not have been strongly weighted when forming expectations of future data due their spatial distance from the most recent data. Indeed, when other attentional cues are in closer spatial proximity to the task at hand, arrows are thought to be less effective in automatically directing visual attention (Leblanc & Jolicoeur, 2010). A third possibility is that the trend line is explicit, and therefore does not require extra processing. In contrast, the arrow requires mental extrapolation, which presumably takes additional cognitive effort.

In summary, simply cueing cognition for the slope of the global trend using an arrow was not effective in influencing expectations for future data. Data across studies 7 and 8 are consistent with expectations being anchored on the most recent information available in a time-series graph. (i.e. the data and/or a trend line, when present).

Study 9: Expectations in horizontal and vertical planes

If expectations are anchored on the most recent data and adjusted to account for global trend direction, then emphasizing the global trend direction may increase the weighting of it when making adjustments. In this next study, the orientation of the plotted data (horizontal or vertical) is investigated to see to what extent this might influence expectations by facilitating comprehension of global trends.

Time-series graphs are typically plotted with time on the horizontal x-axis, moving from left-to-right. The horizontal plane, as opposed to the vertical plane, is the predominant plane in which we experience motion and is readily mapped to representations of time (Tversky, 2011). Although the language one speaks may influence spatial conceptualization of time through linguistic metaphors (Boroditsky, 2001; Fuhrman, et al. 2011), mappings between space and time are remarkably flexible in relation to spatial frames of reference (Torralbo, Santiago, & Lupiáñez, 2006). For example, geologic time can be conceptualized in a vertical plane with past-future mapped from bottom-to-top, consistent with the layering of rock strata over time (Kastens & Ishikawa, 2006).

However, unlike horizontal planes, vertical planes have an inherent bias in directionality due to gravity (Tversky, 2011). Furthermore, mental simulations of static images are influenced by representational momentum (Freyd, 1983), including momentum caused by implied gravity in which objects are mentally animated downwards in the gravitational plane (Freyd, Pantzer, & Cheng, 1988; Hubbard, 1997; Hubbard & Ruppel, 2000). Rotating graphs such that past-future time is represented spatially as moving top-to-bottom, in line with the gravitational plane, might therefore generally facilitate perceptual simulation of past data into the future.

Further, representational momentum effects have been found to be mediated by the plane in which the direction of implied motion is occurring (see Hubbard, 2015 for a recent review). Specifically, Hubbard and Bharucha (1988) and Hubbard (1990) report larger effects of representational momentum in the horizontal plane compared to vertical plane. Hence, there might be a weakening

of the impact of recent runs on expectations in vertical planes compared with the horizontal planes.

To explore this possibility, Study 9 followed the same design as Studies 7 and 8, but manipulated the orientation of graph to be either horizontal (with time running left-to-right) or vertical (with time running top-to-bottom), in place of a trend line/arrow manipulation.

Method

Design

To test expectations for the plotted data-series, the same forced choice task as Study 7 was employed. Graphs were either orientated horizontally with the time-points travelling left-right, or vertically with the time-points travelling top-bottom (Figure 30). As with Study 7 and 8, the global trend direction (positive or negative) and recent data (recent-consistent, recent-up, or recent-down) were manipulated. The experiment was therefore a 2 (Orientation) x 2 (Global Trend Direction) x 3 (Recent Data) design, with all variables within participants.

Participants

Forty undergraduate students (35 female, 5 male) with normal or corrected to normal vision, from the University of East Anglia took part in the study in return for course credit or a nominal payment. Average age was 21 years (range 18-57 years).

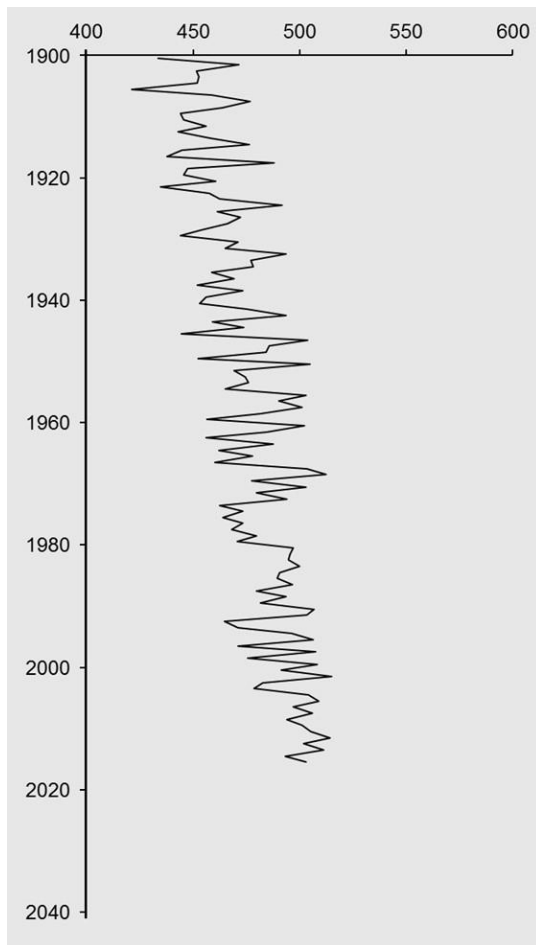


Figure 30. Example of vertical graph stimuli.

Apparatus

Apparatus was the same as Study 7. Vertical graph trials were presented on portrait monitors (same monitors used for horizontal graphs trials, but rotated through 90° with resolution 720 x 1280 pixels). Yes/no responses for judgements about the graphs were mapped to either the 'Z' and 'M' keyboard keys (horizontal response mapping) or to the 'I' and 'M' keyboard keys (vertical response mapping). Keys and horizontal/vertical mappings were reversed and counterbalanced across participants. Confidence ratings using a visual analogue scale (VAS) were controlled using the mouse as before.

Graph Stimuli and Graph Probes

Graph stimuli for horizontal graph trials were the same as the trend line absent stimuli used in Study 7. Graphs for vertical graph trials were identical to those horizontal trials, except they were rotated clockwise through 90°. The text of the y-axis label was orientated horizontally, consistent with Experiment 1. Graph probes were calculated as Study 7 and 8; for horizontal graph trials the probes varied in location along the vertical plane to the right of the screen, while for vertical graph trials the probes varied in location along the horizontal plane at the bottom of the screen.

Procedure

The procedure was the same as Study 7, except that the two blocks of trials consisted of horizontal trials and vertical trials.

Results

Screening criteria was as per Study 7, resulting in 1.53% of trials removed from further analysis. Response distributions were calculated for each cell of the experimental design as per Study 7.

Changes in mean location of expected future values

The mean scores of the response distributions across conditions were submitted to a 2 (Orientation [horizontal, vertical]) x 2 (Global Trend Direction [positive, negative]) x 3 (Recent Data [recent-consistent, recent-up, recent-down]) x 2 (Block Order [horizontal-vertical, vertical-horizontal]) mixed ANOVA (Table 18). Means and standard deviations for each cell of the analysis are provided in Appendix 8, Table A8-1.

There was a main effect of orientation, where there was a significant difference between a negative bias observed for horizontal graphs ($M = -0.103$, $SD = 0.351$) and a positive bias observed for vertical graphs ($M = 0.141$, $SD = 0.383$); $d = 0.664$, 95% CI [0.280, 1.042]. There was a main effect of direction, where there was a positive bias for positive graphs ($M = 0.585$, $SD = 0.520$) and a negative bias for negative graphs ($M = -0.547$, $SD = 0.726$); $d = 1.794$, 95% CI

[1.113, 2.460]. There was also a main effect of recent-data, where there was a greater positive bias for recent-up data ($M = 0.809$, $SD = 0.479$) and a greater negative bias for recent-down data ($M = -0.779$, $SD = 0.331$), when compared with recent-consistent data ($M = 0.027$, $SD = 0.373$); $t(39) = 12.150$, $p < .001$, $d = 1.821$ [1.318, 2.315], $t(39) = 12.074$, $p < .001$, $d = 2.287$ [1.666, 2.926], respectively.

There was also a main effect of block order, where there was a significant difference between a positive bias observed for participants who received horizontal trials first ($M = 0.158$, $SD = 0.313$), and a negative bias observed for participants who received vertical trials first ($M = -0.120$, $SD = 0.222$); $t(38) = 3.232$, $p = .003$, $d = 1.022$, 95% CI [0.355, 1.677]. The effect was moderated by a global trend x block order two-way interaction, in which there was greater negative bias for graphs with negative global trends when receiving vertical graph trials first ($M = -0.976$, $SD = 0.540$), than when receiving horizontal graph trials first ($M = -0.119$, $SD = 0.635$), $t(38) = 4.599$, $p < .001$, $d = 1.454$ [0.747, 2.147]. Conversely the difference between the positive bias for graphs with positive global trends between block orders was not significant ($M = 0.737$, $SD = 0.429$; $M = 0.434$, $SD = 0.568$), $t(38) = 1.900$, $p = .065$, $d = -0.601$, [-1.231, 0.037]. There was no interaction between orientation and global trend direction (Figure 31) or between orientation and recent data (Figure 32).

Table 18. Study 9 mixed ANOVA for changes in mean location of expected future values; * indicates significance at the .05 level.

Source	Test	<i>p</i>	ηp^2
Main effects			
orientation	$F(1,38) = 14.441$.001*	.275
global trend direction	$F(1,38) = 56.609$	< .001*	.598
recent data	$F(2,76) = 245.547$	< .001*	.866
block order	$F(1,38) = 10.447$.003*	.216
Two-way interactions			
orientation x global trend direction	$F(1,38) = 1.562$.219	.039
orientation x recent data	$F(2,76) = 0.974$.382	.025
orientation x block order	$F(1,38) = 2.990$.092	.073
global trend direction x recent data	$F(2,76) = 2.957$.058	.072
global trend direction x block order	$F(1,38) = 14.832$	<.001*	.281
recent data x block order	$F(2,76) = 0.594$.554	.015
Three-way interactions			
orientation x global trend direction x recent data	$F(2,76) = 0.438$.647	.011
orientation x global trend direction x block order	$F(1,38) = 1.344$.254	.034
orientation x recent data x block order	$F(2,76) = 2.352$.102	.058
global trend direction x recent data x block order	$F(2,76) = 0.858$.428	.022
Four-way interaction			
orientation x global trend direction x recent data x block order	$F(2,76) = 0.822$.443	.021

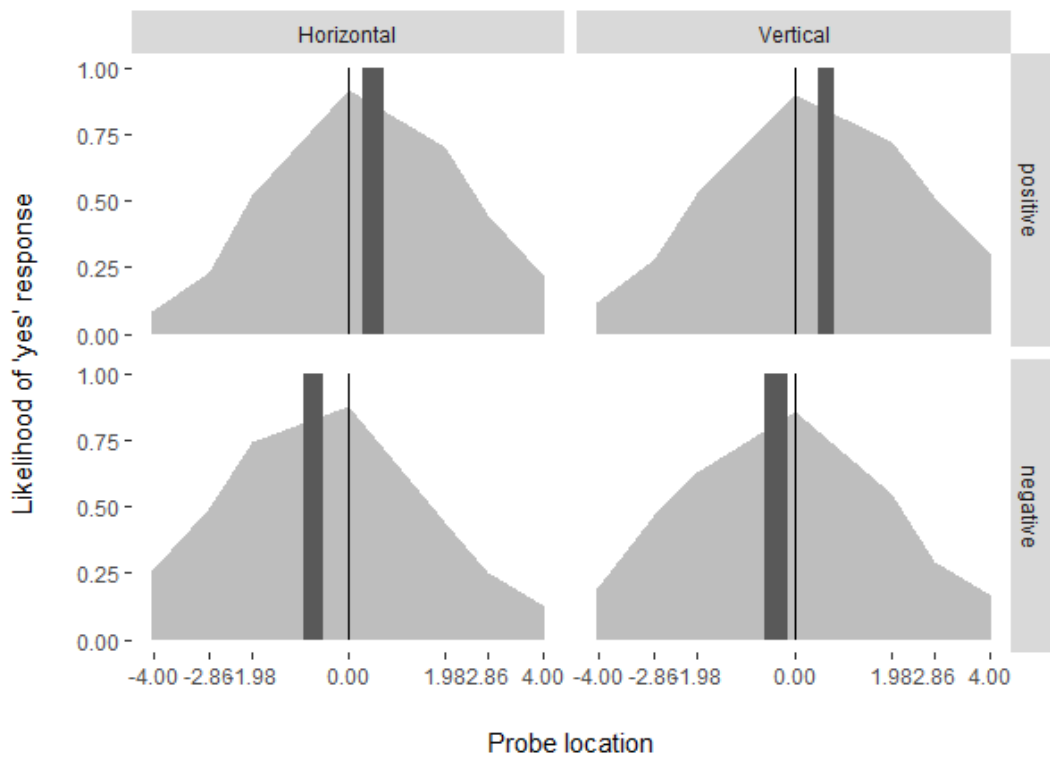


Figure 31. No interaction between orientation and global trend direction. Vertical dark grey bars indicate the 95% confidence interval for the mean location of expectation distributions. Vertical line at probe location 0 provided as a reference point. Light grey shaded areas indicate the full distribution of 'yes' responses for each condition.

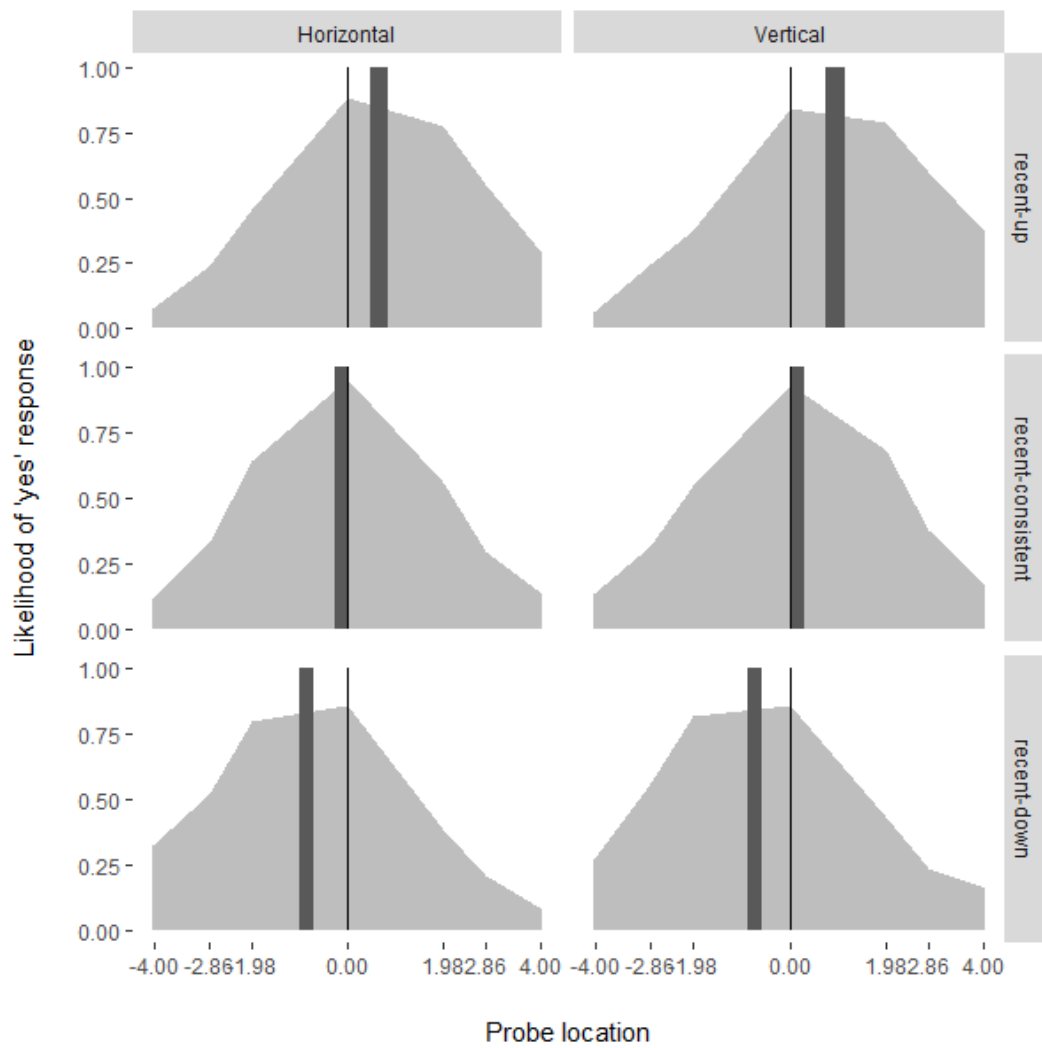


Figure 32. No interaction between orientation and recent data. Vertical dark grey bars indicate the 95% confidence interval for the mean location of expectation distributions. Vertical line at probe location 0 provided as a reference point. Light grey shaded areas indicate the full distribution of 'yes' responses for each condition.

Expectations for future values outside 95% range

Mean ‘yes’ responses to the probes outside the 95% range were compared by submitting scores to a 2 (Orientation [horizontal, vertical]) x 2 (Block Order [horizontal-vertical, vertical-horizontal]) mixed ANOVA. Means and standard deviations for each cell of the analysis are provided in Appendix 8, Table A8-2. There was no main effect of orientation, $F(1,38) = 2.618$, $p = .114$, $\eta p^2 = .064$, or of block order, $F(1,38) = 2.007$, $p = .165$, $\eta p^2 = .050$, and no orientation x block order interaction, $F(1,38) = 0.003$, $p = .958$, $\eta p^2 < .001$.

Confidence in judgements

Mean VAS scores for trials in which the probe was judged consistent with past data were submitted to a 2 (Orientation [horizontal, vertical]) x 2 (Probe location [central, outer]) x 2 (Block Order [horizontal-vertical, vertical-horizontal]) mixed ANOVA (Table 19). Means and standard deviations for each cell of the analysis are provided in Appendix 8, Table A8-3.

There was a main effect of probe location, in which confidence scores were greater for central probes than outer probes ($M = 69.616$, $SD = 12.384$; $M = 59.677$, $SD = 10.679$, respectively; $t(39) = 8.682$, $p < .001$, $d = 0.860$, 95% CI [0.585, 1.128]). No other main effects or interactions were statistically significant.

Table 19. Study 9 mixed ANOVA for mean confidence ratings on ‘yes’ judgements; * indicates significance at the .05 level.

Source	Test	<i>p</i>	ηp^2
Main effects			
orientation	$F(1,38) = 1.197$.281	.031
probe location	$F(1,38) = 80.970$	<.001*	.681
block order	$F(1,38) = 0.008$.930	<.001
Two-way interactions			
orientation x probe location	$F(1,38) = 0.039$.845	.001
orientation x block order	$F(1,38) = 0.583$.450	.015
probe location x block order	$F(1,38) = 3.896$.056	.093
Three-way interaction			
orientation x probe location x block order	$F(1,37) = 1.092$.303	.028

Discussion

Consistent with Study 7 and 8, expectations for horizontal graphs showed trend anti-dampening in response to global trends in the data, and continuation of runs of recent data, per the hot-hand phenomenon. There was no interaction between graph orientation and global trend direction, nor an interaction between graph orientation and recent data, indicating that aligning the representation of time congruent to the gravitational plane did not influence trend anti-dampening or continuation of runs in recent data.

However, a main effect of orientation indicated a general downward bias in expectations for future values in horizontal graphs, relative to a general positive bias in expectations for future values in vertical graphs (from the perspective of the viewer, these translate to a downward spatial bias and a rightward spatial bias, respectively). The downward spatial bias suggests the possibility of representational gravity acting on expectations, similar to implied gravitational effects on mental simulations of objects depicted in static images (Freyd, Pantzer, & Cheng, 1988; Hubbard, 2005). However, this does not account for the

unexpected positive (rightward) spatial bias for vertical graphs, but there is some evidence to suggest that rightward spatial biases are stronger than leftward spatial biases (Halpern & Kelly, 1993; see Hubbard, 2015 for discussion). The potential causes for the interaction between block order and global trend direction are also unclear (in which there was a greater downward bias for graphs with downward global trends when the block of vertical graphs were presented first, compared to when the block of horizontal graphs were presented first). It is currently unclear why this pattern occurred and further investigation is warranted to see if this effect is replicated.

Given that vertical graphs did not narrow expectations, influence confidence in judgements, or reduce effects of trend anti-dampening or continuation of runs, the evidence points to there being no, or very limited, effects of representational gravity on influencing expectations about future data. Effects of implied gravity on spatial memory errors are comparatively small in magnitude (Ziemkiewicz & Kosara, 2010; Freyd, Pantzer, & Cheng, 1988), whereas stronger spatial memory errors have been observed for implied 'attraction' between visual features in visual displays, especially in contexts in which features conform to Gestalt principles of grouping (Ziemkiewicz & Kosara, 2010). This therefore suggests that implied dynamics in general, rather than gravitational dynamics per se, may influence perceptual simulation of plotted data. In other words, the momentum of direction implied by connected lines may be significantly greater in magnitude than momentum from implied gravity, and therefore account for why graph orientation had no or minimal effects on future expectations.

General discussion

Across studies 7-9, expectations of future data from time-series graphs show trend anti-dampening and continuation of recent runs of data (Table 20). Trend lines added to graphs counteracted trend anti-dampening and continuation of runs. However, no such effects were found for arrows or for vertically orientated graphs with time represented as moving in the direction of gravity. Trend lines narrowed the range of expectations for future data and increased confidence in

expectations for data points congruent with the trend line, and reduced confidence for data points either side of the trend line. Trend lines therefore appear to be powerful visual features that not only summarise the global trend of past data, but also direct our expectations of future data.

Trend lines may be particularly salient, drawing visual attention, and have well understood meaning among graphically literate individuals. As the trend lines ran throughout the length of the plotted data, the spatial region on the graphs that contained the most recent data also included the trend line. If the most recent data is preferentially attended to when forming expectations of the future, then the trend line is likely to also be attended to. Conversely, arrows were smaller in length (and so may have been less visually salient), may have had ambiguous meaning (Tverksy, 2011), and were not located in close spatial proximity to the most recent data. Given that making inferences from features contained within visual displays depends both on bottom-up visual processing and top-down knowledge (Hegarty, Canham, & Fabrikant, 2010; Hegarty, 2011) this may explain why arrows did not influence expectations in the same way trend lines did. Providing readers with knowledge of the meaning and/or relevance of the arrows may therefore enhance their use to inform expectations. Furthermore, providing an explicit visual cue, such as a trend line, enables the reader to offload cognition onto perception; conversely, inferring the trend from an arrow requires spatial processing to extrapolate the trend, requiring greater cognitive effort.

Table 20. Summary of key findings across studies 7-9.

	Study manipulation		
	Study 7: trend line	Study 8: arrow	Study 9: orientation
Trend anti-dampening? (main effect of global trend direction on mean location of expectations)	✓	✓	✓
Manipulation counteracts trend anti-dampening? (manipulation x global trend direction interaction on mean location of expectations)	✓	✗	✗
Recent-data hot-hand effect? (main effect of recent data on mean location of expectations)	✓	✓	✓
Manipulation counteracts recent-data hot-hand effect? (manipulation x recent data interaction on mean location of expectations)	✓	✗	✗
Manipulation narrows expectations? (main effect of manipulation on 'yes' response rates to outer probes)	✓	✗ ⁶	✗
Greater confidence in judgements for central probes than outer probes? (main effect of probe location on confidence ratings of 'yes' responses)	✓	✓	✓
Manipulation increases confidence in judgements for central probes and decreases confidence for outer probes? (manipulation x probe location interaction on confidence ratings of 'yes' responses)	✓	✗	✗

⁶ The main effect was present, but was moderated by a block effect, suggesting a demand characteristic rather than a real effect.

Finding trend anti-dampening and continuation of runs in recent data in vertical graphs, consistent with horizontal graphs, suggests that interpretation of spatial relationships encoded in graphs are highly flexible in nature. Just as time can be conceptualised in multiple spatial frames of reference (Torralbo, Santiago, & Lupiáñez, 2006), so too can quantity (Tverksy, 2011; Tverksy, Kugelmass, & Winer, 1991), together with relationships between the two depicted in two-dimensional space. Further, the non-symmetrical nature of the vertical plane (due to gravity), relative to the symmetrical horizontal plane (Tverksy, 2011), does not appear to meaningfully influence inferences.

The observed anti-trend dampening observed across the current studies is consistent with effects for one step-ahead forecasts for data-series with similarly shallow trends (Harvey & Reimers, 2013). Continuation of runs in recent data away from global trends in the absence of trend lines, is also consistent with one-step ahead forecasting research (Lawrence & O'Connor, 1992). As hypothesized by Harvey & Reimers (2013), forecasts may be selected from a range of (uncertain) forecasts, with ecological knowledge of patterns of trends influencing where in this range a forecast is made. The current set of studies add credence to this possibility by confirming that expectations about future data are indeed uncertain and conceptualised across a range of possibilities with varying confidence. Further, as with point forecasts, ecological knowledge of time-series data might influence the nature of the uncertainty distributions. For example, if time-series experienced in the environment typically show a high degree of autocorrelation with immediately prior timepoints, then the range of expectations for runs in recent data would be expected to be weighted in the direction of the run.

It is important to note that although trend lines can support expectations for future data consistent with global trends, this may not always be desirable from a communications perspective. In the example of global average temperatures presented in the introduction of this chapter, fitting a trend line to the data might mitigate the extent to which non-experts believe the data indicates a slow-down, or pause, in warming. However, fitting trend lines involves numerous normative and/or potentially subjective decisions, such as over what

time period to fit the line, choice of function (e.g. linear, polynomial, moving average), and for a given function, the choice of estimation method. Such statistical transformations of underlying data may cause unease among non-experts and reduce trust in the data (Walsh, 2014; Walsh, 2015). Furthermore, finding that a trend line narrows the range of expectations for future values is consistent with accounts that individuals may focus on the statistical mean (represented by the trend line) and discount possible future values at the extremes of the statistical distribution (Spiegelhalter, Pearson & Short, 2011; Broad, et al, 2010). Hence, reliance on a trend line might discount consideration of ‘best’ and ‘worst’ case scenarios for the future, which, for example, in the context of climate change data might influence decisions regarding mitigation and adaptation. Decisions of whether to use a trend line or not should therefore consider the context of the data and the communication goal.

Chapter 5: Discussion

Cognitive models regarding comprehension of data visualisations have largely been founded on cognition for simple datasets and well-defined tasks (Hegarty, 2011). However, in real-world contexts, such as the communication of climate change, data visualisations can contain complex information and require the reader to more generally interpret information in order to draw inferences regarding the meaning of the data. For example, in contrast to simple tasks such as reading off values for specific data points, climate change data visualisations may require the reader to interpret patterns in noisy data or make inferences from the data about the future. These tasks suggest a role for a spatial processing (Trickett & Trafton, 2006). However, cognition and spatial inferences for complex data visualisations, particularly in the context of climate change, have received limited empirical investigation. This thesis therefore set out to achieve two aims. First, to understand the goals, contexts and constraints of the IPCC's communication of climate change via data visualisations. Second, to empirically investigate cognition of data visualisations where spatial inferences may be required – namely in time-series graphs, a common format used to communicate how indicators of a changing climate vary over time.

This chapter summarises and synthesizes findings across the studies presented in this thesis, and identifies how cognitive and psychological science insights could support climate change researchers to enhance the accessibility (i.e. the ease of comprehension) of data visualisations to non-expert (i.e. in IPCC communications). The chapter also considers limitations of the research and provides suggestions for future work.

Results overview

To understand the role of data visualisations in IPCC reports that communicate climate science, Chapter 2 presented a thematic analysis of interviews with IPCC authors regarding the figures in the IPCC Fifth Assessment Report, Working Group 1, SPM (Study 1). This analysis identified that a key consideration in the production of the figures is the need to maintain a high level of scientific rigour, which results in the creation of figures that contain a high level of informational complexity. Further, the analysis identifies that authors do not expect policy-makers to be able to understand the figures as presented – rather they expect policy makers to enlist the support of experts to make sense of them. Study 2 demonstrated that authors generally have a good awareness of which type of figures non-experts might perceive as being difficult to comprehend. Experts' (climate scientists) rankings of the ten Working Group 1 SPM figures for their expected ease of comprehension by non-experts aligned with the ranking provided by non-experts (university undergraduates) for their perceived ease of comprehension. Figures that non-experts considered to be more difficult to comprehend were associated with higher degrees of visual complexity (Study 3), aligning with IPCC authors' beliefs that visually complex figures will be more difficult for non-experts to understand (Study 1). These studies highlight a challenge – how can scientific information be presented in data visualisations such that non-expert audiences can more easily understand them while also maintaining scientific rigour of the presented information?

Here, an understanding of the cognitive processes involved in comprehending complex data visualisations can provide important insights on how this might be achieved. Taking time-series graphs as an example of a common format used to communicate aspects of climate science, Chapter 3 examined encoding of trends from time-series data. Finding in a pilot study (Study 4) that non-experts (university undergraduates) do not always describe trends in time-series graphs that show short-term variability (i.e. noise), Study 5 and Study 6 then investigated to what extent trend information and short-term variability information is encoded into mental representations. Previous work has shown that information not explicitly represented in a data visualisation must be

inferred using spatial processing (Trafton, et al., 2002). Identifying trends within noisy data requires the trend to be spatially inferred from the data and therefore requires cognitive resources and effort (Freedman & Shah, 2002; Carswell, Emery, & Lonon, 1993). Given that language can support spatial processing, and IPCC authors indicated the important role of linguistic explanations to support non-experts' understanding of IPCC figures (Study 1), it is of particular interest to understand how language might support spatial representations of trends.

Results from Studies 5 and 6 demonstrate that a linguistic warning, alerting a reader to ignore extreme data in time-series graphs and attend to trends, improved mental spatial representations of trends. This evidence lends support for a spatial processing component in cognitive models of comprehension of data visualisations (Tricket & Trafton, 2006). In addition, Study 5 demonstrated that the succinct warning acted directly on visual attention. When studying the graphs, those that received the warning spent longer fixating on the area of the graph consistent with the long-term trend. This evidence is consistent with the interaction between bottom-up perceptual processes and top-down knowledge during comprehension of data visualisations (Pinker, 1990; Shah & Freedman, 2002, Hegarty, 2011), and indicates that language (top-down knowledge) can support spatial inferences by acting on perceptual processes. Interestingly, Study 6, found no evidence to suggest that a warning to support representation of short-term variability conferred any benefit, unlike the warning to support representations of long-term trends. The attentional advantages of warnings (and language more generally) may therefore be particularly beneficial in contexts requiring spatial inferences, such as spatially inferring patterns in data, but may have limited benefit to encoding features that are already salient, presumably because bottom-up visual processing is adapted to this task (Itti & Koch, 2001).

Another spatial inference related to time-series graphs is about how a pattern of data will evolve into the future. The primary communication goal of an IPCC SPM is to provide policy-relevant information to support decision-making (Study 1; IPCC, 2016). Inferences made about the future from historic observations of climate data (which are typically plotted in time-series graphs) has been particularly contentious in the context of the so called 'global warming

pause' (Lewandowsky, 2011; Kerr, 2009; Lean & Rind, 2009). Previous research has found that people's forecasts from time-series data (i.e. specific predictions at future time points, typically one-step ahead) deviate from patterns contained in the historic data plotted in the graphs (Bolger & Harvey, 1993; Lawrence & Makridakis, 1989; Harvey & Reimers, 2013). However, interpretation of time-series data in relation to the future does not necessarily involve making specific forecasts; general expectations can also be made covering a spread of possible futures. Studies 7-9 found that expectations for the future data in time-series graphs showed trend anti-dampening, i.e. existing trends were expected to accelerate. These findings are consistent with trend anti-dampening effects found in forecasting studies (Bolger & Harvey, 1993; Lawrence & Makridakis, 1989; Harvey & Reimers, 2013) and support the suggestion that expectations may be uncertain, and may represent a range of possible expectations (Harvey & Reimers, 2013). This set of studies also found that the individuals anchor expectations on the most recent data points in the time-series. Expectations were weighted in the direction of short runs of data away from the long-term trend. Findings are consistent with the hot-hand effect (Gilovich, Vallone & Tversky, 1985) and weighting judgements on recent data in forecasting studies (Lawrence & O'Connor, 1992; Bolger and Harvey, 1993). Recent data had a greater effect on expectations away from the long-term trend than anti-dampening effects. This is consistent with findings in Studies 4-6, in that trend information may not be easily inferred, and so may have only a limited influence on expectations about the future in comparison to more salient features such as recent data. Directly representing the trend in the graph via a trend line counteracted both anti-trend dampening and weighting of expectations in the direction of recent runs (Study 7). Here, not only can spatial inferences be offloaded onto perceptual processes (Hegarty 2011; Trickett & Trafton, 2006), but trend information (i.e. the trend line) is visually salient. It is important to note that a trend line does not *support* spatial inferences about the trend, but rather removes the need to make the spatial inferences.

Study 8 then considered whether an arrow, indicating the direction of the trend, might support spatial inferences for the trend and therefore influence

expectations about the future. Arrows can support spatial processing of static images through mental animations (Heiser & Tversky, 2006). Unlike a trend line, which directly represents the slope of a trend, providing an arrow potentially supports the reader to *infer* the trend's slope using spatial processing. However, arrows did not influence expectations about the future. Trend anti-dampening and continuation of runs of recent data persisted. The lack of an effect of the arrows might be because the meaning of the arrow in this context was not understood (Tversky, 2011). In contrast to trend lines, arrows are not common graphical features in time-series graphs. Alternatively, it might be that the arrows were not attended to during study (i.e. encoding) as they were not perceptually salient. While arrows are thought to automatically capture visual attention in simple visual stimuli (Hommel, et al., 2001), when embedded in a complex data visualisation they may be less salient. Further to this point, given that expectations about the future show anchoring on recent data, which is usually on the right-hand side of graphs, individuals might have largely ignored the left-hand side where the arrows were placed. A further possibility is that interpreting the arrow in the task may require additional cognitive resources, and individuals might simply avoid effortful cognitive processing where possible (Kahneman & Frederick, 2002).

Given that mental animations of static images are known to be influenced by representational gravity (Freyd, Pantzer, & Cheng, 1988; Hubbard, 1997; Hubbard & Ruppel, 2000), Study 9 considered whether the orientation of the graph, horizontal or vertical, influences future expectations. In contrast to the addition of a trend line (Study 7) or an arrow (Study 8), interpreting a graph that is simply rotated through 90° to the vertical does not require additional perceptual processing as there are no added graph features. Vertical graphs, with time running top-to-bottom, might support spatial inferences because the vertical plane is aligned with representational effects of gravity (see Tversky, 2011) in which representational momentum effects tend to be smaller than in horizontal plans (Hubbard & Bharucha, 1988; Hubbard, 1990). Furthermore, spatial memory has been shown to be more accurate for locations below the eye-line compared to above the eye-line (Wilson, et al., 2004; Wilson, et al., 2007). However, vertical

graphs did not influence expectations for the future differently to horizontal graphs – individuals showed trend anti-dampening and continuation of runs in recent data for both horizontal and vertical graphs. Individuals seemed perfectly able to interpret the graphs in the potentially unfamiliar vertical orientation, as expectations were consistent with expectations in the more common horizontal orientation. Hence, implied representational gravity did not seem to affect spatial inferences. It is possible that the stronger effect of representational momentum of the connected in the direction of time (left-to-right, or top-to-bottom), regardless of the orientation of the line, may explain the consistency between horizontal and vertical graphs (Ziemkiewicz & Kosara, 2010).

Evidence from Studies 7-9 suggests that characteristics of data visualizations that act on bottom-up perceptual processes (e.g. arrows and orientation) may be largely ineffective in *supporting* spatial inferences in data visualizations. However, visual characteristics that *replace* spatial inferences (e.g. trend lines) appear to be particularly effective. The pattern of evidence across these studies supports the case for a spatial component in cognitive models of comprehension of data visualizations, in which spatial processing is employed when information is not directly represented in the visualization and therefore has to be inferred (Trickett & Trafton, 2006). However, the current evidence further indicates that when inferences can be made by using spatial processing *or* perceptual processing, perceptual processing wins out. Whether spatial or perceptual processing is employed may of course be dependent on the context in which inferences are made. For example, perceptual processing might be the default approach when individuals make fast heuristic-based inferences, whereas the use of spatial processing may be more likely when analytic processes are employed that override default perceptual processes (see Evans, 2003). This possibility is further supported by evidence indicating that spatial processing requires additional cognitive resources in comparison to perceptual processing (Freedman & Shah, 2002; Carswell, Emery, & Lonon, 1993), and evidence indicating that, when making quick judgements, people tend to be cognitive misers (Kahneman & Frederick, 2002).

Enhancing the comprehension of data visualizations may be achieved by removing the need for spatial inferences by directly representing the relevant information (Trafton, et al., 2000). However, in real-world contexts, this may not always be appropriate. Readers of visualisations may not want simplifications and may prefer more detail (Hegarty, et al., 2009). Further, statistical transformations plotted in data visualisations, such as trend lines, may be interpreted as statistical ‘tricks’, potentially causing a lack of trust among non-experts (Walsh, 2014; Walsh, 2015). In addition, simplifying data visualisations, such that inferences can be drawn using fast intuitive judgements might result in superficial interpretation of the data. As identified in the interviews with IPCC authors, a further concern is that simplifying data visualisations (e.g. in order to remove the need for spatial inferences) may also come with the cost of losing scientific rigour (Study 1). Consequently, in this context, it would be useful to support readers in making spatial inferences, rather than simplifying content to avoid the need for spatial inferences.

As outlined above, *guiding* spatial inferences via visual features that act on bottom-up perceptual processing may be of limited success. Conversely, supporting spatial inferences by providing top-down knowledge may be more promising. As demonstrated in Studies 5 and 6, providing prior knowledge via a linguistic warning supported spatial inferences for trends. Other research has found similar findings in other contexts, for example, providing instructions about how to interpret spatial information in weather maps was more effective than simply adjusting the perceptual salience of task-relevant features (Hegarty, Canham, & Fabrikant, 2010). Data visualisations (i.e. external visual representations) and language both enable symbolic representation of spatial information (Boroditsky, 2001; Coventry, et al., 2010) and drawing on both of these symbolic representations when interpreting spatial information may therefore confer a cognitive advantage.

Here it is of interest to note that the IPCC authors placed emphasis on linguistic explanations to support comprehension of the IPCC Working Group 1 SPM figures (Study 1). IPCC authors acknowledged that many of the figures were visually complex and difficult to understand, but argued that supporting

explanations enabled non-experts to understand them. This may help to explain, to some extent, why the IPCC figures are generally highly regarded by IPCC authors, but their ease of comprehension is criticised by others who are less familiar with their content (IPCC, 2016). It might be that the authors' first-hand experience in communicating the figures, where they are able to provide supporting explanations, is that people are generally able to comprehend them. Conversely, readers may struggle to interpret the information contained in figures when an IPCC author (or other expert familiar with their content) is not on hand to provide supporting explanations. IPCC authors may therefore be unaware of the extent of possible comprehension problems, especially as figures are not empirically tested during the production of reports, and feedback to the drafts of IPCC reports has historically been sought predominately from other experts (IPCC, 2016).

Translating insights from cognitive and psychological research into practice

Given the IPCC's desire to maintain a high degree of scientific rigour in the figures of IPCC reports (Study 1; IPCC, 2016), and the potential for comprehension difficulties among non-experts in understanding the figures (McMahon, Stauffacher, & Knutti, 2015; Studies 2-4), how might the IPCC enhance the accessibility of figures for future reports?

As identified in the introduction (Chapter 1), which reviews psychological and cognitive science evidence, there is the opportunity to draw on evidence-based insights to create figures that are easier for non-experts to comprehend, while maintaining scientific rigour. This goal aligns with the IPCC's current desire to make the output of future reports more accessible and user-friendly to diverse audiences (IPCC, 2016). In addition, improving the ease of accessibility of data visualisations of climate science also has implications for how society might make best use of scientific knowledge. There have been calls for climate scientists to take participatory roles in co-productive frameworks alongside stakeholders to help inform societal decision-making (Rapley, 2014). Data

visualisations of climate data that are accessible to all parties involved could support improved engagement, dialogue and decision-making between scientists, policy-makers, practitioners, communities and publics. Climate service providers (who supply tailored climate knowledge to decision-makers) often use data visualisations to communicate findings, and although the communication goals and intended audience may be much more specific in these contexts than the global assessments made by the IPCC, data visualisation challenges remain (Davis, et al., 2016).

While the science underpinning the comprehension of data visualisations is still developing, general guidelines to support climate scientists in making scientific figures more accessible to non-expert audiences can be drawn from insights from the cognitive science and psychological literature (Chapter 1), together with the work presented in this thesis. Table 21 summarises these insights and provides associated guidelines to improve accessibility of IPCC figures, and indeed data visualisations in general. Guidelines 6 and 11 in Table 20 draw on insights developed from the research presented in this thesis.

Table 21. Evidence-informed guidelines to improve accessibility of scientific data visualisations of climate science.

Psychological insights	Associated guidelines to improve accessibility
<p>1. Intuitions about effective data visualisations do not always correspond to evidence-informed best practice for increasing accessibility (Smallman & St John 2005; Zacks, et al., 1998; Hegarty et al., 2009).</p>	<p>Use cognitive and psychological principles to inform the design of data visualisations; test data visualisations during their development to understand viewers' comprehension of them (McMahon, Stauffacher, & Knutti, 2015; Hegarty, 2011).</p>
Direct visual attention	
<p>2. Visual attention is limited and selective – visual information may or may not be looked at and/or processed by viewers (Simons & Chabris, 1999).</p>	<p>Present only the visual information that is required for the communication goal at hand (Kosslyn, 2006).</p> <p>Direct viewers' visual attention to visual features of the data visualisation that support inferences about the data (Kosslyn, 1989).</p>
<p>3. Salient visual features (where there is contrast in size, shape, colour or motion) can attract visual attention (Wolfe & Horowitz, 2004; Bruce, Green, & Georgeson, 2003).</p>	<p>Make important visual features of the data visualisation perceptually salient so that they 'capture' the attention of the viewer (Kosslyn, 1989).</p>

Table 21 (continued).

Psychological insights	Associated guidelines to improve accessibility
Direct visual attention (continued)	
<p>4. Prior experience and knowledge can direct visual attention (Peebles & Cheng, 2003; Hegarty, Canham, & Fabrikant, 2010).</p>	<p>Choose and design data visualisations informed by viewers' familiarity and knowledge of using visuals and their knowledge of the domain, i.e. knowledge about what the data represents (Kosslyn, 2006).</p> <p>Provide knowledge to viewers about which features of the data visualisation are important to look at, e.g. in text positioned close to the data visualisation (see Guideline 10).</p>
Reduce complexity	
<p>5. An excess of visual information can create visual clutter and impair comprehension (Neider & Zelinsky, 2011; Baldassi, Megna, & Burr, 2006; Coco & Keller, 2009).</p>	<p>Only include information that is needed for the intended purpose of the data visualisation (Kosslyn, 2006); break down the data visualisation into visual 'chunks', each of which should contain enough information for the intended task or message (Shah, Mayer, & Hegarty, 1999).</p>
Support inference-making	
<p>6. Some inferences may require spatial processing of the data (Trafton, et al., 2005, Studies 5-9); experts may have strong spatial reasoning skills (Shipley, et al., 2013), non-experts may not (Hambrick, 2012).</p>	<p>Remove or reduce the need for spatial reasoning skills by showing inferences directly in the data visualisation (Trafton, et al., 2000; Study 7) and/or</p> <p>Support viewers in spatial reasoning, by providing guidance in text (Study 4 and 5). See also Guideline 10.</p>

Table 21 (continued).

Psychological insights	Associated guidelines to improve accessibility
Support inference-making (continued)	
7. The visual structure and layout of the data influences inferences drawn about the data (Shah & Carpenter, 1995).	Identify the most important relationships in the data that are to be communicated; consider different ways of structuring the data that enable the viewer to quickly identify these relationships (Kosslyn, 2006).
8. Animating a data visualisation may help or hinder comprehension (Tverksy, Morrison, & Betrancourt, 2002; Mayer, et al., 2005).	Decisions to create animated data visualisations should be informed by cognitive principles (Shipley, Fabrikant, Lautenschütz, 2013). Consider providing user-control over the playback and speed of the animation (Betrancourt, 2005).
9. Conceptual thought often makes use of cultural metaphors (Lakoff & Johnson, 1980).	Match the visual representation of data to metaphors that aid conceptual thinking, e.g. ‘up’ is associated with ‘good’ and ‘down’ is associated with ‘bad’ (Lakoff & Johnson, 1980); data with negative connotations may be easiest to understand if presented in a downwards direction (Meier & Robinson, 2004).
Integrate text with data visualisations	
10. When the data visualisation and the associated text are spatially distant, attention is split (Mayer, 2009; Holsanova, Holmberg, & Holmqvist, 2009).	Keep the data visualisation and accompanying text close together (Tufte, 2006), e.g. use text within a visual and locate the visual next to the accompanying body text.

Table 21 (continued).

Psychological insights	Associated guidelines to improve accessibility
Integrate text with data visualisations (continued)	
11. Language can influence thought about a visual (see Study 5 and 6; Coventry, et al., 2013).	Use text to help direct viewers' comprehension of the data visualisation, i.e. by providing key knowledge needed to interpret the visual (Kosslyn, 2006).

Putting guidance in to practice

Applying the guidelines to IPCC figures can improve their accessibility to both expert and non-expert audiences. To demonstrate this, the guidelines in Table 21 have been applied to one of the IPCC working Group 1 SPM figures that was perceived by IPCC authors as being challenging to understand (Harold, et al., 2016). Climate change researchers (i.e. experts) and academic researchers from other disciplines (i.e. non-experts) indicated a preference for the revised version of the figure created using the guidelines over the original figure. Such user testing, together with assessing comprehension and cognition can help inform the development of cognitive inspired figures as part of an iterative design cycle.

The guidance presented in Table 21 provides a framework for creating more accessible data visualisations. However, as individuals and groups can differ, there is no substitute for empirically testing data visualisations with the target audience. In the context of IPCC reports, such testing may be seen as an extra burden on an already demanding process (Stocker & Plattner, 2014). However, such testing need not be costly or time-consuming. Asking people to look at and interpret drafts of data visualisations can indicate if data visualisations are broadly understandable or not. Furthermore, rich diagnostic

evidence afforded by eye tracking can indicate the efficiency of comprehension and can identify reasons why comprehension is impaired, such as assessing whether task-relevant information is visually salient or not. Informed by such evidence, appropriate adjustments to data visualisations can be made and they can be re-tested.

Greater collaboration between the climate change research community and the psychology and cognitive science community could help to realise such an approach. For example, as the IPCC looks ahead to their Sixth Assessment Report, there is an opportunity for the IPCC to open-up the review process and ask for feedback on drafts of SPM figures (Harold, et al., 2016). Promisingly, the IPCC have already started to take on-board this suggestion, as the review process for the IPCC's Special Report on 1.5° specifically asks individuals to include comments on communication aspects of the figures (IPCC 2017b). Climate scientists and psychologists could also jointly develop cognitively-inspired visualisations of climate data, that are both accessible and scientifically robust, for use in outputs outside of the formal IPCC process (so-called 'derivative products'). Similar collaborations between research communities have led to improved communication in related fields such as cartography (Fabrikant, Hespanha, & Hegarty, 2010) and geoscience (Shipley, et al., 2013).

Visualisations of climate data are integral to scientific assessments of climate change, but only support communication and decision-making *if* they are understood by the target audience. Empirically testing data visualisations, and applying insights from the science of human cognition to help overcome comprehension problems, offers the potential to make climate science knowledge more accessible to decision-makers in society, while also retaining the integrity of the scientific data and evidence on which they are based.

Limitations

While it is hoped that the research presented in this thesis, together with the proposed guidelines, will have value in supporting climate change researchers

endeavouring to enhance the communication of their findings, it is important to mention potential limitations of the work to date. The research was undertaken with two groups of individuals – climate change scientists (Study 1) and university undergraduate students (Studies 2-9). In particular, Study 2 and Study 4 required students to interpret IPCC figures. However, university students are not the primary audience for these figures. From the perspective of IPCC authors, the primary audience are technical analysts working in government (Study 1). Hence, although this thesis identifies that non-expert audiences may experience difficulties in interpreting some of the IPCC Working Group 1 SPM figures, it is not known to what extent these difficulties are reflective of technical analysts. Judgements made from data visualisations of climate change model outputs have been shown to differ between university students and representatives of governments engaged in international climate change negotiations (Bosetti, et al., 2017). However, audiences of IPCC reports are broader than just technical analysts (IPCC, 2016). Indeed, university undergraduates might be a reasonable proxy for policy-makers outside of an expert technical analyst group, as neither policy makers (in general), nor undergraduate students studying subjects other than climate change, would be expected to hold a high level of scientific expertise about climate change.

Critically, the ease of comprehension of IPCC figures could be enhanced for all audiences. Highly educated audiences from disciplines other than climate science can struggle to interpret IPCC figures (McMahon, Stauffacher, & Knutti, 2015) and spatial processing abilities can differ between experts coming from different scientific specialisms (Resnick & Shipley, 2013). Given that IPCC reports bring together research from across the natural and social sciences, it is important that figures are accessible to experts with different domain expertise. In this sense, if a figure can be comprehended by undergraduate students then there is a good chance it can be comprehended by other educated individuals in society and a broad range of experts.

The research presented in this thesis only considers the figures created by the IPCC Working Group 1, which covers the physical science basis. The figures created by IPCC Working Group 2 (impacts, adaptation and vulnerability) and

Working Group 3 (mitigation of climate change) are not considered. While there may be differences in the approach that authors take to data visualisation in Working Group 2 and 3 compared to Working Group 1, all three working groups use scientific figures to support communication and there is a desire to improve the ease of comprehension of scientific figures across all three working groups (IPCC, 2016). Furthermore, the guidelines provided in Table 21 are not tied to data visualisations in specific domains, but rather are general enough to be applied to a wide-range of subject domains.

Another limitation of the research is that the insights generated in experimental settings using controlled stimuli (i.e. studies 5-9) have not yet been validated in more ecologically valid settings. In these studies, very limited context about the time-series graphs was provided to control for differences in prior knowledge across participants. Importantly, however, the time-series graphs represented complex data sets. Cognitive models regarding the comprehension of data visualisations have largely been informed by experiments using simple datasets (Hegarty, 2011). Hence, the current studies therefore contribute evidence in support of scaling-up models to data visualisations that are more representative of those used in real-world settings.

Related to the above limitation, figures in IPCC reports are accompanied by supporting linguistic information in the form of figure captions and text in the main body of reports. Given that language can support spatial inferences (Study 5 and 6), it might be that captions and text provide additional context, influencing cognition and comprehension. However, text and visual information may not always be read in conjunction with one another, especially when visual and text elements are spatially separated (Holsanova, Holmberg, & Holmqvist, 2009). Furthermore, the nature of supporting text in reports such as IPCC SPMs is largely descriptive, rather than instructive. For example, figure captions typically provide descriptions of what visual features represent, and report text may highlight a message that the figure is intended to convey. However, the text may not provide guidance about how to *read* a figure, i.e. akin to the warning instruction tested in Study 5 and 6. It would therefore be beneficial to further

investigate how different types of linguistic information influence comprehension (see also future directions below).

Future directions

Given that language can support spatial inferences by equipping readers with prior knowledge (Study 5 and 6), other forms of prior knowledge could be considered in future work. For example, prior beliefs have been shown to influence judgements about data presented in data visualisations (Lewandowsky, 2011; Shah, 2002). However, it is not known to what extent prior beliefs act directly on cognitive processes when making spatial inferences with data visualisations. It is possible that prior beliefs about, and/or a high degree of familiarity for, a data visualisation might result in limited cognitive processing of the data visualisation. For example, IPCC figures have been shown to instil a sense of confidence in their scientific integrity following only very brief presentation, whereas more simplistic figures were seen to be less credible, suggesting that people make quick judgements based on their expectations (McMahon, Stauffacher, Knutti, 2016). Hence, prior beliefs and expectations might influence the extent of cognitive effort exerted when interpreting data visualisations, determining the extent to which spatial inferences are made. Such effects might be contextualised in relation to dual-process theories of cognition (Evans & Stanovich, 2013).

Given that IPCC authors identified linguistic explanations as being important to facilitate understanding of data visualisations (Study 1), and language can support spatial inferences (Study 5 and 6), it would be of interest to evaluate the type of linguistic explanations used by IPCC authors when explaining the figures. As mentioned above, the content of figure captions and text might be very different to verbal explanations when guiding a reader's understanding of a figure (for example when IPCC authors explain the figures face-to-face with an audience). This could help evaluate the extent to which authors support readers' comprehension through instructive language (e.g. saying "look at the trend", akin to the warning in Study 5 and 6) and evaluate to what extent such descriptions

support spatial inferences more generally across the range of figures used by the IPCC.

In relation to the guidelines presented in Table 21, there is a need to evaluate how climate change researchers, such as IPCC authors, might best be able to apply the guidance in practice. For example, breaking up complex information into visual chunks might require extra page space, which might not be possible if there are restrictions on the layout and length of reports. The guidelines presented in Table 21 are currently being adapted to a more practical format, accompanied with visual examples, to encourage their implementation by the IPCC and the climate change community in general. Furthermore, ongoing dialogue with the IPCC regarding some of the work contained in this thesis presents an opportunity for collaborating with IPCC authors to further test out, refine and extend the guidance. For example, climate change uncertainties can be challenging to visually communicate (McMahon Stauffacher, Knutti, 2015) and insights from collaborative work between climate scientists and psychologists could enable tailored guidance to be developed to help overcome such challenges.

Conclusions

There are four conclusions that can be drawn from the work contained in this thesis:

First, IPCC authors are aware that their SPM figures are visually complex and that they may be difficult for non-expert audiences to comprehend. The IPCC is keen to make future reports, including figures, more accessible to such audiences. However, the challenge faced is that the figures are required to be scientifically rigorous, which is perceived as a significant constraint on how greater accessibility might be achieved.

Second, and further to the first point, the cognitive and psychological sciences can offer key insights into how data visualisations could be made easier to understand while maintaining scientific rigour. For example, the studies

presented in this thesis demonstrate that when information from a figure must be inferred using spatial processes (i.e. because the information is not explicitly represented), cognition can be supported by providing top-down knowledge, i.e. via language. In contrast, visual features aimed at supporting spatial inferences by acting on bottom-up perceptual processing may be less effective.

Third, evidence from the studies in this thesis supports the need to include a spatial processing component in models of cognition of data visualisations.

Fourth, the guidelines presented in this thesis provide a framework for the application of cognitive and psychological insights to the design and communication of complex data visualisations. Collaboration between psychologists and climate change researchers in applying and advancing the guidelines could provide an opportunity to not only support communication, but also to advance theoretical knowledge in ecologically valid contexts.

References

- Abramowitz, J. S., Tolin, D. F., & Street, G. P. (2001). Paradoxical effects of thought suppression: A meta-analysis of controlled studies. *Clinical Psychology Review, 21*(5), 683-703.
- Alexander, A. L., Kaber, D. B., Kim, S. H., Stelzer, E. M., Kaufmann, K., & Prinzel III, L. J. (2012). Measurement and modeling of display clutter in advanced flight deck technologies. *The International Journal of Aviation Psychology, 22*(4), 299-318.
- Andrienko, G., Andrienko, N., Demsar, U., Dransch, D., Dykes, J., Fabrikant, S.I., Jern, M., Kraak, M.J., Schumann, H. and Tominski, C. (2010). Space, time and visual analytics. *International Journal of Geographical Information Science, 24*(10), 1577-1600.
- Atance, C. M., & O'Neill, D. K. (2001). Episodic future thinking. *Trends in Cognitive Sciences, 5*(12), 533-539.
- Ayton, P., & Fischer, I. (2004). The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness?. *Memory & Cognition, 32*(8), 1369-1378.
- Baldassi, S., Megna, N., & Burr, D. C. (2006). Visual clutter causes high-magnitude errors. *PLoS Biology, 4*(3): e56.
- Barkemeyer, R., Dessai, S., Monge-Sanz, B., Renzi, B. G., & Napolitano, G. (2016). Linguistic analysis of IPCC summaries for policymakers and associated coverage. *Nature Climate Change, 6*, 311-317.
- Baylis, G. C., & Driver, J. (1995). One-sided edge assignment in vision: 1. Figure-ground segmentation and attention to objects. *Current Directions in Psychological Science, 4*(5), 140-146.

- Betrancourt, M. The animation and interactivity principles in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 287-296). Cambridge: Cambridge University Press.
- Black, R. (2015). No more summaries for wonks. *Nature Climate Change*, 5(4), 282.
- Bolger, F., & Harvey, N. (1993). Context-sensitive heuristics in statistical reasoning. *The Quarterly Journal of Experimental Psychology*, 46(4), 779-811.
- Borkin, M. A., Gajos, K. Z., Peters, A., Mitsouras, D., Melchionna, S., Rybicki, F. J., Fedlman, C. L., & Pfister, H. (2011). Evaluation of artery visualizations for heart disease diagnosis. *IEEE Transactions on Visualization and Computer Graphics*, 17(12), 2479-2488.
- Borkin, M. A., Vo, A. A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., & Pfister, H. (2013). What makes a visualization memorable?. *IEEE Transactions on Visualization and Computer Graphics*, 19(12), 2306-2315.
- Boroditsky, L. (2001). Does language shape thought?: Mandarin and English speakers' conceptions of time. *Cognitive Psychology*, 43(1), 1-22.
- Bosetti, V., Weber, E., Berger, L., Budescu, D. V., Liu, N., & Tavoni, M. (2017). COP21 climate negotiators' responses to climate model forecasts. *Nature Climate Change*, 7(3), 185-189.
- Boykoff, M. T. (2014). Media discourse on the climate slowdown. *Nature Climate Change*, 4, 156-158.
- Boykoff, M. T., & Boykoff, J. M. (2007). Climate change and journalistic norms: A case-study of US mass-media coverage. *Geoforum*, 38(6), 1190-1204.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101.

- Broad, K., Leiserowitz, A., Weinkle, J., & Steketee, M. (2007). Misinterpretations of the “cone of uncertainty” in Florida during the 2004 hurricane season. *Bulletin of the American Meteorological Society*, 88(5), 651-667.
- Bruce, V., Green, P. R., & Georgeson, M. A. (2003). *Visual perception: physiology, psychology, & ecology*. New York: Psychology Press.
- Brunyé, T. T., & Taylor, H. A. (2009). When goals constrain: Eye movements and memory for goal-oriented map study. *Applied Cognitive Psychology*, 23(6), 772-787.
- Budescu, D. V., Broomell, S., & Por, H. H. (2009). Improving communication of uncertainty in the reports of the Intergovernmental Panel on Climate Change. *Psychological Science* 20(3), 299-308.
- Budescu, D. V., Por, H. H., Broomell, S. B., & Smithson, M. (2014). The interpretation of IPCC probabilistic statements around the world. *Nature Climate Change*, 4, 508-512.
- Burnham, P., Lutz, K. G., Grant, W., & Layton-Henry, Z. (2008). *Research methods in politics*. New York: Palgrave Macmillan.
- Burns, B. D., & Corpus, B. (2004). Randomness and inductions from streaks: “Gambler’s fallacy” versus “hot hand”. *Psychonomic Bulletin & Review*, 11(1), 179-184.
- Camerer, C., Loewenstein, G., & Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. *Journal of Political Economy*, 97(5), 1232-1254.
- 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., Emery, C., & Lonon, A. M. (1993). Stimulus complexity and information integration in the spontaneous interpretations of line graphs. *Applied Cognitive Psychology*, 7(4), 341-357.
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, 4, 55-81.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36, 181-204.
- Clarke, V., Braun, V., & Hayfield, N. (2015). Thematic Analysis. In J. A. Smith (Ed), *Qualitative psychology: A practical guide to research methods* (pp. 222-248). London: Sage.
- Coco, M. I., & Keller, F. The impact of visual information on reference assignment in sentence production. In N. Taatgen, & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp.274-279).
- Cohen, C. A., & Hegarty, M. (2007). Individual differences in use of external visualisations to perform an internal visualisation task. *Applied Cognitive Psychology*, 21, 701-711.
- Correll, M., Albers, D., Franconeri, S., & Gleicher, M. (2012). Comparing averages in time series data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1095-1104), Austin Texas. ACM.
- Coventry, K. R., Christophel, T. B., Fehr, T., Valdés-Conroy, B., & Herrmann, M. (2013). Multiple routes to mental animation: Language and functional relations drive motion processing for static Images. *Psychological Science*, 24(8), 1379-1388.
- Coventry, K. R., Lynott, D., Cangelosi, A., Monrouxe, L., Joyce, D., & Richardson, D. C. (2010). Spatial language, visual attention, and perceptual simulation. *Brain and Language*, 112(3), 202-213.

- Crampton, J. W. (2002). Interactivity types in geographic visualization. *Cartography and Geographic Information Science*, 29, 85-98.
- Cumming, G. (2014). *Understanding the new statistics: Effect sizes confidence intervals, and meta-analysis*. New York: Routledge.
- Da Silva, M. P., Courboulay, V., & Estraillier, P. (2011). Image complexity measure based on visual attention. In *(ICIP, 2011) 18th IEEE International Conference on Image Processing* (pp. 3281-3284).
- Daron, J. D., Lorenz, S., Wolski, P., Blamey, R. C., & Jack, C. (2015). Interpreting climate data visualisations to inform adaptation decisions. *Climate Risk Management*, 10, 17-26.
- Davis, M., Lowe, R., Steffen, S., Doblus-Reyes, F. & Rodó, X. (2016). Barriers to using climate information: Challenges in communicating probabilistic forecasts to decision-makers. In J. L. Drake, Y. Y. Kontar, J. C. Eichelberger, T. S. Rupp, & K. M. Taylor (Eds.), *Communicating climate-change and natural hazard risk and cultivating resilience* (pp. 95-113). London: Springer.
- Decety, J., & Grèzes, J. (2006). The power of simulation: imagining one's own and other's behavior. *Brain Research*, 1079(1), 4-14.
- diSessa, A. A. (2004). Metarepresentation: Native competence and targets for instruction. *Cognition and Instruction*, 22(3), 293-331.
- Doyon-Poulin, P., Robert, J. M., & Ouellette, B. (2012). Review of visual clutter and its effects on pilot performance: A new look at past research. In *2012 IEEE/AIAA 31st Digital Avionics Systems Conference (DASC)*, (pp. 2D1-1).
- Driver, J., & Baylis, G. C. (1995). One-sided edge assignment in vision: 2. Part decomposition, shape description, and attention to objects. *Current Directions in Psychological Science*, 4(6), 201-206.

- Eiriksdottir, E., & Catrambone, R. (2011). Procedural instructions, principles, and examples how to structure instructions for procedural tasks to enhance performance, learning, and transfer. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 53(6), 749-770.
- Elzer, S., Carberry, S., Chester, D., Demir, S., Green, N., Zukerman, I., & Trnka, K. (2005). Exploring and exploiting the limited utility of captions in recognizing intention in information graphics. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics* (pp. 223-230). Association for Computational Linguistics.
- Evans, J. S. B. (2003). In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences*, 7(10), 454-459.
- Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, 8(3), 223-241.
- Ewles, H. (2013, October 18). Sir Mark Walport delivers CSap lecture on climate change. University of Cambridge Centre for Science and Policy. Retrieved from <http://www.csap.cam.ac.uk/news/article-mark-walport-csap-lecture-on-climate-change/>
- Fabrikant, S. I., Hespanha, S. R., & Hegarty, M. (2010). Cognitively inspired and perceptually salient graphic displays for efficient spatial inference making. *Annals of the Association of American Geographers*, 100, 13-29.
- Feist, M. I., & Gentner, D. (2007). Spatial language influences memory for spatial scenes. *Memory & Cognition*, 35(2), 283-296.
- Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, 1(3), 288-299.

- Fisher, B., Green, T. M., & Arias-Hernández, R. (2011). Visual analytics as a translational cognitive science. *Topics in Cognitive Science*, 3(3), 609-625.
- Fisher, R. (1955). Statistical methods and scientific induction. *Journal of the Royal Statistical Society. Series B (Methodological)*, 17(1), 69-78.
- 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. 18-30). Berlin: Springer.
- Freyd, J. J. (1983). The mental representation of movement when static stimuli are viewed. *Attention, Perception, & Psychophysics*, 33(6), 575-581.
- Freyd, J. J., & Pantzer, T. M. (1995). Static patterns moving in the mind. In S. M. Smith, T. B. Ward, & R. A. Finke (Eds.), *The creative cognition approach* (pp. 181-204). Cambridge, MA: The MIT Press.
- Freyd, J. J., Pantzer, T. M., & Cheng, J. L. (1988). Representing statics as forces in equilibrium. *Journal of Experimental Psychology: General*, 117(4), 395-407.
- Friesen, C. K., Ristic, J., & Kingstone, A. (2004). Attentional effects of counterpredictive gaze and arrow cues. *Journal of Experimental Psychology: Human Perception and Performance*, 30(2), 319-329.
- Fuhrman, O., McCormick, K., Chen, E., Jiang, H., Shu, D., Mao, S., & Boroditsky, L. (2011). How linguistic and cultural forces shape conceptions of time: English and Mandarin time in 3D. *Cognitive Science*, 35(7), 1305-1328.
- Gentner, D., & Gentner, D. R. (1983). Flowing waters or teeming crowds: Mental models of electricity. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 99-129). Hillsdale, NJ: Erlbaum.

- Gibson, B. S., & Bryant, T. A. (2005). Variation in cue duration reveals top-down modulation of involuntary orienting to uninformative symbolic cues. *Attention, Perception, & Psychophysics*, *67*(5), 749-758.
- Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*, 451-482.
- Gigerenzer, G., Hertwig, R., van den Broek, E., Fasolo, B., & Katsikopoulos, K. V. (2005). "A 30% chance of rain tomorrow": How does the public understand probabilistic weather forecasts? *Risk Analysis*, *25*(3), 623-629.
- Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, *17*(3), 295-314.
- Ginns, P. (2006). Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction*, *16*(6), 511-525.
- Gobet, F. (2005). Chunking models of expertise: Implications for education. *Applied Cognitive Psychology*, *19*(2), 183-204.
- Grant, E. R. & Spivey, M. J. (2003). Eye movements and problem solving: Guiding attention guides thought. *Psychological Science*, *14*(5), 462-466.
- Griffin, A. L., MacEachren, A. M., Hardisty, F., Steiner, E., & Li, B. (2006). A comparison of animated maps with static small-multiple maps for visually identifying space-time clusters. *Annals of the Association of American Geographers*, *96*(4), 740-753.
- Grisson, S., Tipper, S. P., & Hewitt, O. (2005). Long-term negative priming: Support for retrieval of prior attentional processes. *The Quarterly Journal of Experimental Psychology Section A*, *58*(7), 1199-1224.

- Grison, S., Paul, M. A., Kessler, K., & Tipper, S. P. (2005). Inhibition of object identity in inhibition of return: Implications for encoding and retrieving inhibitory processes. *Psychonomic Bulletin & Review*, *12*(3), 553-558.
- Halpern, A. R., & Kelly, M. H. (1993). Memory biases in left versus right implied motion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 471-484.
- Hambrick, D. Z., Libarkin, J. C., Petcovic, H. L., Baker, K. M., Elkins, J., Callahan, C. N., Turner, S. P., Rench, T. A., & LaDue, N. D. (2012). A test of the circumvention-of-limits hypothesis in scientific problem solving: The case of geological bedrock mapping. *Journal of Experimental Psychology: General*, *141*(3), 397-403.
- Harold, J., Lorenzoni, I., Shipley, T. F., & Coventry, K. R. (2016). Cognitive and psychological science insights to improve climate change data visualization. *Nature Climate Change*, *6*(12), 1080-1089.
- Harrower, M., & Brewer, C. A. (2003). ColorBrewer.org: An online tool for selecting colour schemes for maps. *The Cartographic Journal*, *40*(1), 27-37.
- Harvey, N., & Reimers, S. (2013). Trend damping: Under-adjustment, experimental artifact, or adaptation to features of the natural environment? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(2), 589-607.
- Hausfather, Z., Cowtan, K., Clarke, D. C., Jacobs, P., Richardson, M., & Rohde, R. (2017). Assessing recent warming using instrumentally homogeneous sea surface temperature records. *Science Advances*, *3*(1), e1601207.
- Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d' . *Behavior Research Methods, Instruments, & Computers*, *27*(1), 46-51.

- Hawkins, E., Edwards, T., & McNeill, D. (2014). Pause for thought. *Nature Climate Change*, 4(3), 154-156.
- Heaps, C., & Handel, S. (1999). Similarity and features of natural textures. *Journal of Experimental Psychology: Human Perception and Performance*, 25(2), 299-320.
- Hegarty, M. (1992). Mental animation: Inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(5), 1084-1102.
- Hegarty, M. (2011). The cognitive science of visual-spatial displays: Implications for design. *Topics in Cognitive Science* 3(3), 446-474.
- Hegarty, M., Canham, M. S., & Fabrikant, S. I. (2010). Thinking about the weather: How display salience and knowledge affect performance in a graphic inference task. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 36(1), 37-53.
- Hegarty, M., Kriz, S., & Cate, C. (2003). The roles of mental animations and external animations in understanding mechanical systems. *Cognition and Instruction*, 21(4), 325-360.
- Hegarty, M., & Sims, V. K. (1994). Individual differences in mental animation during mechanical reasoning. *Memory & Cognition*, 22(4), 411-430.
- Hegarty, M., Smallman, H. S., Stull, A. T., & Canham, M. S. (2009). Naïve cartography: How intuitions about display configuration can hurt performance. *Cartographica*, 44(3), 171-186.
- Heiser, J., & Tversky, B. (2006). Arrows in comprehending and producing mechanical diagrams. *Cognitive Science*, 30(3), 581-592.
- Henderson, J. M., Weeks, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing.

Journal of Experimental Psychology: Human Perception and Performance, 25(1), 210–228.

- Hermer-Vazquez, L., Spelke, E. S., & Katsnelson, A. S. (1999). Sources of flexibility in human cognition: Dual-task studies of space and language. *Cognitive Psychology*, 39(1), 3-36.
- Heylighen F. (1999). The growth of structural and functional complexity during evolution. In F. Heylighen, J. Bollen & A. Riegler (Eds.), *The evolution of complexity* (pp. 17-44). Dordrecht: Kluwer Academic
- Hinds, P. J. (1999). The curse of expertise: The effects of expertise and debiasing methods on prediction of novice performance. *Journal of Experimental Psychology: Applied*, 5(2), 205-221.
- Ho, H. N., Van Doorn, G. H., Kawabe, T., Watanabe, J., & Spence, C. (2014). Colour-temperature correspondences: When reactions to thermal stimuli are influenced by colour. *PloS ONE*, 9, e91854.
- Hoffman, D. D., & Richards, W. A. (1984). Parts of recognition. *Cognition*, 18(1), 65-96.
- Hoffman, D. D., & Singh, M. (1997). Saliency of visual parts. *Cognition*, 63(1), 29-78.
- Hollin, G. J. S., & Pearce, W. (2015). Tension between scientific certainty and meaning complicates communication of IPCC reports. *Nature Climate Change*, 5, 753-756.
- Holsanova, J., Holmberg, N., & Holmqvist, K. (2009). Reading information graphics: The role of spatial contiguity and dual attentional guidance. *Applied Cognitive Psychology*, 23(9), 1215-1226.
- Hommel, B., Pratt, J., Colzato, L., & Godijn, R. (2001). Symbolic control of visual attention. *Psychological Science*, 12(5), 360-365.

- Hubbard, T. L. (1990). Cognitive representation of linear motion: Possible direction and gravity effects in judged displacement. *Memory & Cognition*, *18*(3), 299-309.
- Hubbard, T. L. (1997). Target size and displacement along the axis of implied gravitational attraction: Effects of implied weight and evidence of representational gravity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(6), 1484-1493.
- Hubbard, T. L. (2005). Representational momentum and related displacements in spatial memory: A review of the findings. *Psychonomic Bulletin & Review*, *12*(5), 822-851.
- Hubbard, T. L. (2015). Forms of momentum across time: behavioral and psychological. *The Journal of Mind and Behavior*, *36*(1/2), 47-82
- Hubbard, T. L., & Bharucha, J. J. (1988). Judged displacement in apparent vertical and horizontal motion. *Perception & Psychophysics*, *44*(3), 211-221.
- Hubbard, T. L., & Ruppel, S. E. (2000). Spatial memory averaging, the landmark attraction effect, and representational gravity. *Psychological Research*, *64*(1), 41-55.
- Huber, O., Wider, R., & Huber, O. W. (1997). Active information search and complete information presentation in naturalistic risky decision tasks. *Acta Psychologica*, *95*(1), 15-29.
- IPCC (2007). Summary for Policymakers. In: S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, & H. L. Miller (Eds.), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK and New York, USA: Cambridge University Press.

- IPCC (2012). Statement about the 2007 Nobel Peace Prize. Retrieved from http://www.ipcc.ch/pdf/nobel/Nobel_statement_final.pdf
- IPCC (2013a). Summary for Policymakers. In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex & P. M. Midgley (Eds.), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK and New York, USA: Cambridge University Press.
- IPCC (2013b). IPCC Factsheet: What is the IPCC? Retrieved from https://www.ipcc.ch/news_and_events/docs/factsheets/FS_what_ipcc.pdf
- IPCC (2014a). Summary for Policymakers. In R. K. Pachauri, & L.A. Meyer (Eds.), *Climate Change 2014: Synthesis Report*. Cambridge, UK and New York, USA: Cambridge University Press.
- IPCC (2014b). Summary for Policymakers. In C. B. Field, V. R. Barros, D. J. Dokken, K. J. Mach, M. D. Mastrandrea, T. E. Bilir, M. Chatterjee, K. L. Ebi, Y. O. Estrada, R. C. Genova, B. Girma, E. S. Kissel, A. N. Levy, S. MacCracken, P. R. Mastrandrea, & L. L. White (Eds.), *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK and New York, USA: Cambridge University Press.
- IPCC (2014c). Summary for Policymakers. In O. Edenhofer, R. Pichs-Madruga, Y. Sokona, E. Farahani, S. Kadner, K. Seyboth, A. Adler, I. Baum, S. Brunner, P. Eickemeier, B. Kriemann, J. Savolainen, S. Schlömer, C. von Stechow, T. Zwickel and J.C. Minx (Eds.), *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK and New York, USA: Cambridge University Press.

IPCC (2016). *Meeting Report of the Intergovernmental Panel on Climate Change Expert Meeting on Communication*. J. Lynn, M. Araya, Ø. Christophersen, I. El Gizouli, S. J. Hassol, E. M. Konstantinidis, K. J. Mach, L. A. Meyer, K. Tanabe, M. Tignor, R. Tshikalanke, J.-P. van Ypersele (Eds.). Geneva, Switzerland: World Meteorological Organization.

IPCC (2017a). The IPCC and the Sixth Assessment Cycle. Retrieved from http://www.ipcc.ch/pdf/ar6_material/AC6_brochure_en.pdf

IPCC (2017b) Expert Review of the IPCC Special Report on 1.5°C Global Warming. Retrieved from http://www.ipcc.ch/report/sr15/pdf/information_note_expert_review.pdf

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2(3), 194-203.

Jacobs, P., Cutting, H., Lewandowsky, S., O'Brien, M., Rice, K., & Verheggen, B. (2015). Clarity of meaning in IPCC press conference. *Nature Climate Change*, 5, 961-962.

Jebb, A. T., Tay, L., Wang, W., & Huang, Q. (2015). Time series analysis for psychological research: examining and forecasting change. *Frontiers in Psychology*, 6, 727.

Joshi, M., Hawkins, E., Sutton, R., Lowe, J., & Frame, D. (2011). Projections of when temperature change will exceed 2 °C above pre-industrial levels. *Nature Climate Change*, 1, 407-412.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgement. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). Cambridge: Cambridge University Press.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3(3), 430-454.

- Karl, T. R., Arguez, A., Huang, B., Lawrimore, J. H., McMahon, J. R., Menne, M. J., Peterson, T. C., Vose, R. S., & Zhang, H. M. (2015). Possible artifacts of data biases in the recent global surface warming hiatus. *Science*, *348*(6242), 1469-1472.
- Kastens, K. A., & Ishikawa, T. (2006). Spatial thinking in the geosciences and cognitive sciences: A cross-disciplinary look at the intersection of the two fields. In C. A. Manduca, & D. W. Mogk (Eds.), *Earth and mind: How geologists think and learn about the Earth* (pp. 53-76). Boulder, CO: GSA Books.
- Kerr, R. A. (2009). What happened to global warming? Scientists say just wait a bit. *Science*, *326*(5949), 28-29.
- Kosslyn, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology*, *3*(3), 185-226.
- Kosslyn, S. M. (2006). *Graph design for the eye and mind*. Oxford: OUP.
- Kövecses, Z. (2005). *Metaphor in culture: Universality and variation*. Cambridge: Cambridge University Press.
- Krasting, J. P., Broccoli, A. J., Dixon, K. W., & Lanzante, J. R. (2013). Future changes in northern hemisphere snowfall. *Journal of Climate*, *26*, 7813-7828.
- Kuhn, G., & Kingstone, A. (2009). Look away! Eyes and arrows engage oculomotor responses automatically. *Attention, Perception, & Psychophysics*, *71*(2), 314-327.
- Lakoff, G., & Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive Science*, *4*(2), 195-208.
- Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, *11*(1), 65-100.

- Lawrence, M., & Makridakis, S. (1989). Factors affecting judgmental forecasts and confidence intervals. *Organizational Behavior and Human Decision Processes*, 43(2), 172-187.
- Lawrence, M., & O'Connor, M. (1992). Exploring judgemental forecasting. *International Journal of Forecasting*, 8(1), 15-26.
- Lean, J. L., & Rind, D. H. (2009). How will Earth's surface temperature change in future decades? *Geophysical Research Letters*, 36(15).
- Learmonth, A. E., Newcombe, N. S., Sheridan, N., & Jones, M. (2008). Why size counts: Children's spatial reorientation in large and small enclosures. *Developmental Science*, 11(3), 414-426.
- Leblanc, É., & Jolicoeur, P. (2010). How do selected arrows guide visuospatial attention? Dissociating symbolic value and spatial proximity. *Journal of Experimental Psychology: Human Perception and Performance*, 36(5), 1314-1320.
- LeFevre, J., & Dixon, P. (1986). Do written instructions need examples? *Cognition and Instruction*, 3(1), 1-30.
- Lewandowsky, S. (2011). Popular consensus: Climate change is set to continue. *Psychological Science*, 22(4), 460-463.
- Light, A., & Bartlein, P. J. (2004). The end of the rainbow? Color schemes for improved data graphics. *Eos*, 85(40), 385-391.
- Loewenstein, J., & Gentner, D. (2005). Relational language and the development of relational mapping. *Cognitive Psychology*, 50(4), 315-353.
- Lowe, R. K. (1999). Extracting information from an animation during complex visual learning. *European Journal of Psychology of Education*, 14(2), 225-244.

- Lowe, R. K. (2003). Animation and learning: Selective processing of information in dynamic graphics. *Learning and Instruction, 13*(2), 157-176.
- Mahony, M., & Hulme, M. (2012). The colour of risk: An exploration of the IPCC's "burning embers" diagram. *Spontaneous Generations: Journal for the History and Philosophy of Science, 6*(1), 75-89.
- Mangos, P. M., & Steele-Johnson, D. (2001). The role of subjective task complexity in goal orientation, self-efficacy, and performance relations. *Human Performance, 14*(2), 169-185.
- Mayer, R.E. (2009). *Multimedia Learning*. Cambridge: Cambridge University Press.
- Mayer, R. E., Hegarty, M., Mayer, S., & Campbell, J. (2005). When static media promote active learning: Annotated illustrations versus narrated animations in multimedia instruction. *Journal of Experimental Psychology: Applied, 11*(4), 256-265.
- McMahon, R., Stauffacher, M., & Knutti, R. (2015). The unseen uncertainties in climate change: reviewing comprehension of an IPCC scenario graph. *Climatic Change, 133*(2), 141-154.
- Meier, B. P., & Robinson, M. D. (2004). Why the sunny side is up: Associations between affect and vertical position. *Psychological Science, 15*(4), 243-247.
- Meyer, J., Shinar, D., & Leiser, D. (1997). Multiple factors that determine performance with tables and graphs. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 39*(2), 268-286.
- Miniukovich, A., & De Angeli, A. (2015). Computation of interface aesthetics. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 1163-1172). ACM.

- Minx, J. C., Callaghan, M., Lamb, W. F., Garard, J., & Edenhofer, O. (2017). Learning about climate change solutions in the IPCC and beyond. *Environmental Science & Policy*, 77, 252-259.
- Moacdieh, N., & Sarter, N. (2015). Display clutter: A review of definitions and measurement techniques. *Human Factors*, 57(1), 61-100.
- Moreland, K. (2009). Diverging Color Maps for Scientific Visualization. In *Proceedings of the 5th International Symposium on Advances in Visual Computing: Part II* (pp. 92-103). Berlin: Springer-Verlag.
- Neider, M. B., & Zelinsky, G. J. (2011). Cutting through the clutter: Searching for targets in evolving complex scenes. *Journal of Vision*, 11, 7.
- Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology*. San Francisco: W. H. Freeman & Company.
- Newman T. P. (2017). Tracking the release of IPCC AR5 on Twitter: Users, comments, and sources following the release of the Working Group I Summary for Policymakers. *Public Understanding of Science*, 26(7), 815-825.
- Nickerson, R. S. (1999). How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others. *Psychological Bulletin*, 125(6), 737-759.
- Nocke, T., Sterzel, T., Böttinger, M., & Wrobel, M. (2008) Visualization of Climate and Climate Change Data: An Overview. In M. Ehlers, K. Behncke, F. Gerstengarbe, F. Hillen, L. Koppers, L. Stroink, & J. Wächter (Eds), *Digital Earth summit on geoinformatics 2008: Tools for global change research*. (pp. 226-232). Heidelberg: Wichmann.
- O'Neill, S., Williams, H. T. P., Kurz, T., Wiersma, B., & Boykoff, M. (2015). Dominant frames in legacy and social media coverage of the IPCC Fifth Assessment Report. *Nature Climate Change*, 5, 380-385.

- Oliva, A., Mack, M. L., Shrestha, M., & Peeper, A. (2004). Identifying the perceptual dimensions of visual complexity of scenes. In *Proceedings of the Cognitive Science Society* 26. (pp. 1041-1046).
- 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(1), 28-46.
- Peng, W., Ward, M. O., & Rundensteiner, E. A. (2004). Clutter reduction in multi-dimensional data visualization using dimension reordering. In *INFOVIS 2004, IEEE Symposium on Information Visualization, 2004* (pp. 89-96).
- Pinker, S. (1990). A theory of graph comprehension. In R. Freedle (Ed.) *Artificial intelligence and the future of testing* (pp 73-126). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Potter, K., Wilson, A., Bremer, P. T., Williams, D., Doutriaux, C., Pascucci, V., & Johnson, C. R. (2009). Ensemble-vis: A framework for the statistical visualization of ensemble data. In *ICDMW'09, IEEE International Conference on Data Mining Workshops* (pp. 233-240).
- Pratt, J., & Hommel, B. (2003). Symbolic control of visual attention: The role of working memory and attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, 29(5), 835-845.
- Rapley, C. G., De Meyer, K., Carney, J., Clarke, R., Howarth, C., Smith, N., & Lotto, B. (2014). *Time for Change? Climate Science Reconsidered: Report of the UCL Policy Commission on Communicating Climate Science*. London: UCL.
- Ratwani, R. M., & Trafton, J. G. (2008). Shedding light on the graph schema: Perceptual features versus invariant structure. *Psychonomic Bulletin & Review*, 15(4), 757-762.

- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, 62(8), 1457-1506.
- Resnick, I., & Shipley, T. F. (2013). Breaking new ground in the mind: an initial study of mental brittle transformation and mental rigid rotation in science experts. *Cognitive Processing*, 14(2), 143-152.
- Rosenholtz, R., Dorai, A., & Freeman, R. (2011). Do predictions of visual perception aid design? *ACM Transactions on Applied Perception (TAP)*, 8(2), 12.
- Rosenholtz, R., Li, Y., Mansfield, J., & Jin, Z. (2005). Feature congestion: a measure of display clutter. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 761-770). ACM.
- Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision*, 7(2), 17.
- Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2007). Task and context determine where you look. *Journal of Vision*, 7, 16.
- Scaife, M., & Rogers, Y. (1996). External cognition: how do graphical representations work?. *International Journal of Human-Computer Studies*, 45(2), 185-213.
- Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: the prospective brain. *Nature Reviews Neuroscience*, 8(9), 657-661.
- Schmidt, A., Ivanova, A., & Schäfer, M. S. (2013). Media attention for climate change around the world: A comparative analysis of newspaper coverage in 27 countries. *Global Environmental Change*, 23(5), 1233-1248.

- Schoemaker, P. J. H. (2004). Forecasting and scenario planning: The challenges of uncertainty and complexity. In D. J. Koehler, & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp.274-296). Oxford: Blackwell Publishing.
- Shah, P. (2002). Graph comprehension: The role of format, content and individual differences. In M. Anderson, B. Meyer, & P. Olivier (Eds.), *Diagrammatic representation and reasoning* (pp. 173-186). London: Springer.
- Shah, P., & Carpenter, P. A. (1995). Conceptual limitations in comprehending line graphs. *Journal of Experimental Psychology: General*, 124(1), 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(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). Cambridge: Cambridge University Press.
- Shah, P., Mayer, R. E., & Hegarty, M. (1999). Graphs as aids to knowledge construction: Signaling techniques for guiding the process of graph comprehension. *Journal of Educational Psychology*, 91(4), 690-702.
- Shaki, S., Fischer, M. H., & Petrusic W. M. (2009). Reading habits for both words and numbers contribute to the SNARC effect. *Psychonomic Bulletin & Review*, 16(2), 328-331.
- Sherman-Morris, K., Antonelli, K. B., & Williams, C. C. (2015). Measuring the effectiveness of the graphical communication of hurricane storm surge threat. *Weather, Climate, and Society*, 7(1), 69-82.
- Shipley, T. F., Fabrikant, S. I., & Lautenschütz, A. K. (2013). Creating perceptually salient animated displays of spatiotemporal coordination in

- events. In M. Raubal, D. M. Mark, & A. U. Frank (Eds.), *Cognitive and linguistic aspects of geographic space* (pp. 259-270). Berlin: Springer.
- Shipley, T. F., Tikoff, B., Ormand, C., & Manduca, C. (2013). Structural geology practice and learning, from the perspective of cognitive science. *Journal of Structural Geology*, *54*, 72-84.
- Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentive blindness for dynamic events. *Perception* *28*(9), 1059-1074.
- Slough, S. W., McTigue, E. M., Kim, S., & Jennings, S. K. (2010). Science textbooks' use of graphical representation: A descriptive analysis of four sixth grade science texts. *Reading Psychology*, *31*(3), 301-325.
- Smallman, H. S., & St. John, M. (2005). Naïve realism: Misplaced faith in realistic displays. *Ergonomics in Design*, *13*(3), 6-13.
- Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174-215.
- Spiegelhalter, D., Pearson, M., & Short, I. (2011). Visualizing uncertainty about the future. *Science*, *333*, 1393-1400.
- Sterman, J. D., & Sweeney, L. B. (2007). Understanding public complacency about climate change: Adults' mental models of climate change violate conservation of matter. *Climatic Change*, *80*(3), 213-238.
- Stocker, T. F., & Plattner, G. K. (2014). Rethink IPCC reports. *Nature*, *513*(7517), 163-166.
- Stocker, T. F., & Plattner, G. K. (2016). Making use of the IPCC's powerful communication tool. *Nature Climate Change*, *6*(7), 637-638.

- Stofer, K., & Che, X. (2014). Comparing experts and novices on scaffolded data visualizations using eye-tracking. *Journal of Eye Movement Research*, 7(5), 2.
- Suddendorf, T., & Corballis, M. C. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans? *Behavioral and Brain Sciences*, 30(3), 299-313.
- Symon, C. (2013). *Climate Change: Action, Trends and Implications for Business*. Cambridge, UK: University of Cambridge.
- Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, J. R. (2009). Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences*, 106(11), 4567-4570.
- Tipples, J. (2002). Eye gaze is not unique: Automatic orienting in response to uninformative arrows. *Psychonomic Bulletin & Review*, 9(2), 314-318.
- Torralba, A., Oliva, A., Castelano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, 113(4), 766-786.
- Torrallbo, A., Santiago, J., & Lupiáñez, J. (2006). Flexible conceptual projection of time onto spatial frames of reference. *Cognitive Science*, 30(4), 745-757.
- Trafton, J. G., Kirschenbaum, S. S., Tsui, T. L., Miyamoto, R. T., Ballas, J. A., & Raymond, P. D. (2000). Turning pictures into numbers: Extracting and generating information from complex visualizations. *International Journal of Human-Computer Studies*, 53(5), 827-850.
- Trafton, J. G., Marshall, S., Mintz, F., & Trickett, S. B. (2002). In M. Hegarty, B. Meyer, & N. H. Narayanan (Eds.), *Diagrammatic Representation and Inference, Volume 2317 Lecture Notes in Computer Science* (pp. 206-220). Berlin: Springer.

- Trafton, J. G., Trickett, S. B., & Mintz, F. E. (2005). Connecting internal and external representations: Spatial transformations of scientific visualizations. *Foundations of Science*, *10*(1), 89-106.
- Trickett, S. B., & Trafton, J. G. (2006). Toward a comprehensive model of graph comprehension: Making the case for spatial cognition. In D. Barker-Plummer, R. Cox, & N. Swoboda (Eds.), *Diagrammatic Representation and Inference, LNAI 4045* (pp. 286-300). Berlin: Springer.
- Trickett, S. B., Trafton, J. G., Saner, L., & Schunn, C. D. (2007). "I don't know what's going on there": The use of spatial transformations to deal with and resolve uncertainty in complex visualizations. In M. C. Lovett, & P. Shah (Eds.), *Thinking with data* (pp. 65-86). Mahwah: Lawrence Erlbaum.
- Tufte, E. R. (2006). *Beautiful Evidence*. New York: Graphics Press.
- Tversky, B. (1997). Cognitive principles of graphic displays. In *AAAI 1997 Fall Symposium on Reasoning with Diagrammatic Representations* (pp. 116-124).
- Tversky, B. (2005). Functional significance of visuospatial representations. In P. Shah & A. Miyake (Eds.), *Handbook of higher-level visuospatial thinking* (pp. 1-34). Cambridge: Cambridge University Press.
- Tversky, B. (2011). Visualizing thought. *Topics in Cognitive Science*, *3*(3), 499-535.
- Tversky, B., Kugelmass, S., & Winter, A. (1991). Cross-cultural and developmental trends in graphic productions. *Cognitive Psychology*, *23*(4), 515-557.
- Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: can it facilitate? *International Journal of Human-Computer Studies*, *57*(4), 247-262.

- Tyler, C. (2013, December 2). Top 20 things scientists need to know about policy-making. *The Guardian*. Retrieved from <http://www.theguardian.com>
- Tyszka, T., Markiewicz, Ł., Kubińska, E., Gawryluk, K., & Zielonka, P. (2017). A belief in trend reversal requires access to cognitive resources. *Journal of Cognitive Psychology*, *29*(2), 202-216.
- Vaughan, C., & Dessai, S. (2014). Climate services for society: origins, institutional arrangements, and design elements for an evaluation framework. *Wiley Interdisciplinary Reviews: Climate Change*, *5*(5), 587-603.
- Walsh, L. (2014). Tricks, hockey sticks, and the myth of natural inscription: The visual rhetoric of climategate. In B. Schneider & T. Nocke (Eds.), *Image politics of climate change: Visualizations, imaginations, documentations* (pp. 81-104). Bielefeld: transcript Verlag.
- Walsh, L. (2015). The visual rhetoric of climate change. *Wiley Interdisciplinary Reviews: Climate Change*, *6*(4), 361-368.
- Wickens, C. D., & Carswell, C. M. (1995). The proximity compatibility principle: Its psychological foundation and relevance to display design. *Human Factors*, *37*(3), 473-494.
- Wickens, C. D., Nunes, A., Alexander, A. L., & Steelman, K. (2005). *3D navigation and integrated hazard display in advanced avionics: Workload, performance, and situation awareness display clutter issues in SVS design*. Illinois: University of Illinois
- Wiehr, F., Setlur, V., & Joshi, A. (2013). DriveSense: Contextual handling of large-scale route map data for the automobile. In *IEEE International Conference on Big Data* (pp. 87-94). IEEE.
- Wilkening, J., & Fabrikant, S. (2011). How do decision time and realism affect map-based decision making?. In *Spatial Information Theory, LNCS 6899, 10th International Conference on Spatial Information Theory* (pp. 1-19).

- Wogalter, M. S., Godfrey, S. S., Fontenelle, G. A., Desaulniers, D. R., Rothstein, P., & Laughery, K. R. (1987). Effectiveness of warnings. *Human Factors*, 29(5), 599-612.
- Wolfe, J. M., & Bennett, S. C. (1997). Preattentive object files: Shapeless bundles of basic features. *Vision Research*, 37(1), 25-43.
- Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, 5(6), 495-501.
- Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum Press.
- Yeh, M., & Wickens, C. D. (2001). Attentional filtering in the design of electronic map displays: A comparison of color-coding, intensity coding, and decluttering techniques. *Human Factors*, 43(4), 543-562.
- Yohe, G., & Oppenheimer, M. (2011). Evaluation, characterization, and communication of uncertainty by the intergovernmental panel on climate change—an introductory essay. *Climatic Change*, 108, 629.
- Zacks, J., Levy, E., Tversky, B., & Schiano, D. J. (1998). Reading bar graphs: Effects of extraneous depth cues and graphical context. *Journal of Experimental Psychology: Applied*, 4(2), 119-138.
- Zacks, J., Levy, E., Tversky, B., & Schiano, D. J. (2002). Graphs in print. In M. Anderson, B. Meyer, & P. Olivier (Eds.), *Diagrammatic representation and reasoning* (pp. 187-206). London: Springer.
- Zacks, J., & Tversky, B. (1999). Bars and lines: A study of graphic communication. *Memory & Cognition*, 27(6), 1073-1079.
- Ziemkiewicz, C., & Kosara, R. (2010). Laws of attraction: From perceptual forces to conceptual similarity. *IEEE Transactions on Visualization and Computer Graphics*, 16(6), 1009-1016.

Appendix 1

Interview protocol / guide

Intro / housekeeping:

- Thank you for agreeing to participate; information sheet
- I'm interested in how experts and novices understand visual displays, such as scientific Figures about climate change, and my research hopes to help inform how visuals might be adapted for different audiences. With particular interest in the work of the IPCC.
- The interview will last no more than 1 hour, and along with the interview questions there will be some ranking tasks, where I'll ask you to order the 10 Figures from the SPM based on different criteria.
- The interview will be recorded and interview transcribed
- All information you provide will be treated as confidential - your name will only be known to the research team, and will not appear on any of the final reports of the research
- Right to withdraw – during or after the interview
- Any questions?
- Consent

Rank task 1 – non-expert audience

- Before exploring with you a policy maker audience, I'd like to start asking you to consider how a non-expert lay audience might interpret the Figures from the SPM
- Figures of the SPM: please spread out the cards in a random order
- I'd first like you to consider how easy or difficult they are to understand by *non-expert audiences*
- Please rank order the Figures from the one you think – university undergraduates without climate science training - would find easiest to

understand (rank 1) through to the one that you think they would find the most difficult to understand (rank 10)

- Collect up cards

Work with the IPCC

- Thank you – now I'd like to ask you about your expertise and role with the IPCC
- Prompts
 - How many years have you worked with the IPCC?
 - What roles have you held over that time?
 - Which reports have you authored?
 - What were your roles and responsibilities for the AR5 Working Group 1?
 - Can you tell me about your role in authoring the Working Group 1 Summary for Policy Makers?

Overview of involvement in each of the Figures

- Spread out the cards again
- First, I'd like to ascertain which, if any, of the Figures you had involvement in
- Which of the Figures did you have:
 - Involvement in the collection or analysis of all or part of the data that makes up the Figure?
 - Input in to the design and/or creation of the Figure ?
 - Reviewed and/or commented on the Figure during the creation of the report ?
- Thank you – I'll come back to your involvement in those figures a bit later

Audiences: Who is the intended audience for SPM Figures?

- The report is titled as being for policy makers....

- Are there particular types of policy makers that the Figures were created for?
 - Who are they? / What are their characteristics?
 - Why were they created for those groups?
- When the report was being created, were the figures *aimed at* any other audiences?
- Do you think the Figures would be useful to other audiences? If so, who?

Rank task 2 – policy makers

- [Shuffle cards]
- I'd like you now to consider the Figures from the perspective of policy makers
- Please rank order the Figures from the one you think policy makers would find easiest to understand (rank 1) through to the one that you think they would find the most difficult to understand (rank 10).
- [Shuffle cards]

Purpose, involvement and process

- What do you think is the main purpose of including Figures in the report)?
- You mentioned you were involved in the [...] of Figures [...]
- [Was your involvement fairly consistent across these Figures, or was it different?]
- Prompts
 - Can you talk me through your involvement for one of the Figures? (Which Figure? Why?)
 - What were your roles and responsibilities?
 - What was the Figure trying to communicate/achieve?
 - Can you tell me about the processes involved in creating the Figure for the SPM? (choosing / drafting / designing / refining / finalising)
 - Where did the Figure come from originally?

- Who else was involved in the process for the Figures? (at the different stages?)
- [If more than one Figure] What were the differences in process?
- If not involved in design/creation: What is your perception of the process for the creation of the Figures for the SPM? How do you think it works?

Process: Strengths and weaknesses

- What do you think are the strengths of the process in which the Figures are generated?
- Do you think there are any challenges or difficulties with the process? What are these?

Figures: Feedback, strengths and weaknesses

- Following the publication of the report, what feedback have you received from policy makers about the Figures?
- From other audiences ?
 - Have they understood them? (who?)
 - Have any of the Figures caused confusion or disagreement?
- To what extent do you think the Figures achieve their intended purpose?
- What do you feel are the strengths of the Figures?
 - Do you think any are particularly well designed? Which ones? Why?
- From your perspective and experience, do you think there any limitations or problems with any of the Figures?
 - Which ones?
 - What are limitation/problems?
- Are there any ways in which you think any of the Figures could be improved?

Use (by the IPCC and beyond)

- In what ways have the Figures been used in IPCC activities beyond the report?
- Are certain figures from the SPM used and referred to more than others?
 - Which figures, by whom, in what context?
- How have other people/organisations used or adapted the figures (and associated messages)?
 - Who?
 - For what audiences?
 - Were they adapted or used as is?
 - In what ways were they adapted?

Rank task 3 – importance

- [Shuffle cards]
- Now in the final ranking task, I'd like you to rank order the Figures based on their importance to help inform future climate policy, from the one you think is the most important (rank 1) through to the one that you think is least important (rank 10).
- For the most important, why do you think this is the most important?

Closing

- Anything else you would like to add?
- Debrief and outline next steps
- Thank you for your time

Appendix 2

Graph stimuli allocation to blocks for Study 5.

Table A2-1.

Study 5 same-different graph stimuli allocation to blocks.

Dataset	Trend direction	Test graph condition	X-ticks	Block
1	upward	same	2	1
1	upward	same	5	2
1	upward	same	9	3
2	upward	gradient different	2	3
2	upward	gradient different	5	1
2	upward	gradient different	9	2
3	upward	amplitude different	2	2
3	upward	amplitude different	5	3
3	upward	amplitude different	9	1
4	upward	completely different	2	1
4	upward	completely different	5	2
4	upward	completely different	9	3
5	downward	same	2	3
5	downward	same	5	1
5	downward	same	9	2
6	downward	gradient different	2	2
6	downward	gradient different	5	3
6	downward	gradient different	9	1
7	downward	amplitude different	2	1
7	downward	amplitude different	5	2
7	downward	amplitude different	9	3
8	downward	completely different	2	3
8	downward	completely different	5	1
8	downward	completely different	9	2
9	flat	same	2	2
9	flat	same	5	3
9	flat	same	9	1
10	flat	gradient different	2	1
10	flat	gradient different	5	2
10	flat	gradient different	9	3
11	flat	amplitude different	2	3
11	flat	amplitude different	5	1
11	flat	amplitude different	9	2
12	flat	completely different	2	2
12	flat	completely different	5	3
12	flat	completely different	9	1

Table A2-2.

Study 5 filler trials graph stimuli allocation to blocks.

Dataset	Trend direction	Test graph condition	X-ticks	Block
13	upward	describe	2	1
13	upward	describe	5	2
13	upward	describe	9	3
14	downward	describe	2	3
14	downward	describe	5	1
14	downward	describe	9	2
15	flat	describe	2	2
15	flat	describe	5	3
15	flat	describe	9	1
16	upward	comprehension	2	1
17	upward	comprehension	5	2
18	upward	comprehension	9	3
19	downward	comprehension	2	3
20	downward	comprehension	5	1
21	downward	comprehension	9	2
22	flat	comprehension	2	2
23	flat	comprehension	5	3
24	flat	comprehension	9	1

Appendix 3

Data tables (means and standard deviations) for Study 5 analyses.

Table A3-1.

Study 5 means and standard deviations of sensitivity (d') as a function of test graph, warning and x-ticks.

Condition	Test graph					
	Amplitude different		Gradient different		Completely different	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
No warning						
2 x-ticks	0.792	0.798	0.343	0.558	1.357	0.614
5 x-ticks	0.819	0.554	0.276	0.858	1.546	0.426
9 x-ticks	0.537	0.972	0.220	0.875	1.405	0.639
Warning						
2 x-ticks	0.396	1.029	0.415	0.914	1.461	0.660
5 x-ticks	0.359	1.008	0.422	1.002	1.157	0.684
9 x-ticks	0.804	0.781	0.784	0.790	1.537	0.583

Note: No warning, $n = 18$; Warning, $n = 16$.

Table A3-2.

Study 5 means and standard deviations of sensitivity (d') as a function of test graph, warning and block.

Condition	Test graph					
	Amplitude different		Gradient different		Completely different	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
No warning						
Block 1	0.911	0.986	0.566	0.874	1.378	0.705
Block 2	0.592	0.943	0.212	0.503	1.396	0.585
Block 3	0.661	1.014	0.071	0.747	1.519	0.695
Warning						
Block 1	0.532	1.296	0.653	0.782	1.278	0.710
Block 2	0.495	1.005	0.497	0.837	1.463	0.648
Block 3	0.532	0.982	0.427	1.042	1.423	0.785

Note: No warning, $n = 18$; Warning, $n = 16$.

Appendix 4

Graph stimuli allocation to blocks for Study 6.

Table A4-1.

Study 6 same-different graph stimuli allocation to blocks.

Dataset	Trend direction	Test graph condition	Block
1	upward	same	1
2	upward	same	2
3	upward	same	3
4	upward	same	1
5	upward	gradient different	1
6	upward	gradient different	2
7	upward	gradient different	3
8	upward	gradient different	2
9	upward	amplitude different	1
10	upward	amplitude different	2
11	upward	amplitude different	3
12	upward	amplitude different	3
13	upward	completely different	1
14	upward	completely different	2
15	upward	completely different	3
16	upward	completely different	1
17	downward	same	1
18	downward	same	2
19	downward	same	3
20	downward	same	2
21	downward	gradient different	1
22	downward	gradient different	2
23	downward	gradient different	3
24	downward	gradient different	3
25	downward	amplitude different	1
26	downward	amplitude different	2
27	downward	amplitude different	3
28	downward	amplitude different	1
29	downward	completely different	1
30	downward	completely different	2
31	downward	completely different	3
32	downward	completely different	2
33	flat	same	1
34	flat	same	2
35	flat	same	3
36	flat	same	3

Table continued overleaf.

Table A4-1 (continued).

Dataset	Trend direction	Test graph condition	Block
37	flat	gradient different	1
38	flat	gradient different	2
39	flat	gradient different	3
40	flat	gradient different	1
41	flat	amplitude different	1
42	flat	amplitude different	2
43	flat	amplitude different	3
44	flat	amplitude different	2
45	flat	completely different	1
46	flat	completely different	2
47	flat	completely different	3
48	flat	completely different	3

Table A4-2.

Study 6 filler trials graph stimuli allocation to blocks.

Dataset	Trend direction	Test graph condition	Block
49	upward	comprehension	1
50	upward	comprehension	2
51	upward	comprehension	3
52	upward	comprehension	1
53	upward	comprehension	2
54	upward	comprehension	3
55	upward	comprehension	1
56	upward	comprehension	2
57	downward	comprehension	1
58	downward	comprehension	2
59	downward	comprehension	3
60	downward	comprehension	1
61	downward	comprehension	2
62	downward	comprehension	3
63	downward	comprehension	3
64	downward	comprehension	1
65	flat	comprehension	1
66	flat	comprehension	2
67	flat	comprehension	3
68	flat	comprehension	1
69	flat	comprehension	2
70	flat	comprehension	3
71	flat	comprehension	2
72	flat	comprehension	3

Appendix 5

Data tables (means and standard deviations) for Study 6 analyses.

Table A5-1.

Study 6 means and standard deviations of sensitivity (d'), as a function of test graph and block, for the 'no warning' and 'identify trend, ignore extreme warning' groups.

Condition	Test graph					
	Amplitude different		Gradient different		Completely different	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
No warning						
Block 1	1.068	0.861	0.662	0.844	1.553	0.854
Block 2	1.077	0.795	1.031	0.789	1.751	0.591
Identify trend, ignore extreme warning						
Block 1	1.121	0.791	1.175	0.841	1.529	0.781
Block 2	0.960	0.781	1.134	0.938	1.745	0.644

Note: No warning, $n = 18$; Trend goal, ignore and identify warning, $n = 18$.

Table A5-2.

Study 6 means and standard deviations of sensitivity (d') on completely different trials as a function of block, warning user goal and warning informational content.

Condition	Block			
	1		2	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Long-term trend goal				
Ignore and identify	1.529	0.781	1.745	0.644
Ignore only	1.613	0.664	1.620	0.498
Extreme data goal				
Ignore and identify	1.560	0.593	1.553	0.846
Ignore only	1.653	0.680	1.798	0.712

Note: Trend goal, ignore and identify, $n = 18$; Trend goal, ignore only, $n = 18$; Extreme goal, ignore and identify, $n = 18$, Extreme goal, ignore only, $n = 18$.

Table A5-3.

Study 6 means and standard deviations of sensitivity (d') as a function of test graph, warning user goal and warning informational content.

Condition	Test graph					
	Amplitude different		Gradient different		Completely different	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Long-term trend goal						
Ignore and identify	1.116	0.758	1.254	0.767	1.860	0.626
Ignore only	0.820	0.617	0.722	0.756	1.759	0.505
Extreme data goal						
Ignore and identify	0.639	0.809	0.792	0.609	1.699	0.846
Ignore only	1.010	0.775	0.705	0.562	1.971	0.700

Note: Trend goal, ignore and identify, $n = 18$; Trend goal, ignore only, $n = 18$; Extreme goal, ignore and identify, $n = 18$, Extreme goal, ignore only, $n = 18$.

Table A5-4.

Study 6 means and standard deviations of sensitivity (d') as a function of test graph and warning informational content, including the 'no warning' group.

Condition	Test graph					
	Amplitude different		Gradient different		Completely different	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
No warning	1.151	0.695	0.919	0.642	1.840	0.764
Ignore and identify	0.878	0.810	1.023	0.722	1.779	0.738
Ignore only	0.915	0.697	0.713	0.657	1.865	0.611

Note: No warning, $n = 18$; Ignore and identify, $n = 36$; Ignore only, $n = 36$.

Appendix 6

Data tables (means and standard deviations) for Study 7 analyses.

Table A6-1.

Study 7 means and standard deviations of mean location of expected future values (in units of SDs of the distribution of the noise) as a function of recent data direction, trend line, global trend direction and block order.

Condition	Recent data direction					
	Recent-consistent		Recent-up		Recent-down	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Trend line absent						
Negative global trend						
Trend line absent-present	-0.786	0.834	0.339	0.773	-1.434	0.906
Trend line present-absent	-0.499	0.873	-0.172	0.778	-1.262	0.744
Positive global trend						
Trend line absent-present	0.465	0.847	1.488	0.728	-0.236	0.920
Trend line present-absent	0.268	0.901	0.744	1.150	-0.232	0.798
Trend line present						
Negative global trend						
Trend line absent-present	-0.204	0.683	-0.102	0.736	-0.214	0.777
Trend line present-absent	-0.034	0.704	0.069	0.679	-0.415	0.747
Positive global trend						
Trend line absent-present	0.290	0.734	0.330	0.725	-0.039	0.589
Trend line present-absent	0.175	0.853	0.381	0.782	0.092	0.800

Note: NoTrend-Trend, $n = 20$; Trend-NoTrend, $n = 20$.

Table A6-2.

Study 7 means and standard deviations of the mean number of ‘yes’ responses to the outer probes as a function of trend line and block order.

Condition	<i>M</i>	<i>SD</i>
Trend line absent		
Trend line absent-present	0.418	0.105
Trend line present-absent	0.403	0.130
Trend line present		
Trend line absent-present	0.209	0.132
Trend line present-absent	0.303	0.122

Note: NoTrend-Trend, $n = 20$; Trend-NoTrend, $n = 20$.

Table A6-3.

Study 7 means and standard deviations of VAS scores for trials in which the probe was judged consistent with past data, as a function of probe location, trend line and block order.

Condition	Probe location			
	Central		Outer	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Trend line absent				
Trend line absent-present	70.940	10.321	59.657	13.952
Trend line present-absent	78.259	14.876	64.246	15.467
Trend line present				
Trend line absent-present	88.552	11.538	55.813	21.030
Trend line present-absent	85.772	12.897	54.530	11.248

Note: NoTrend-Trend, $n = 19$ (one participant has missing data for one cell of the study design and was not included in the ANOVA); Trend-NoTrend, $n = 20$.

Appendix 7

Data tables (means and standard deviations) for Study 8 analyses.

Table A7-1.

Study 8 means and standard deviations of mean location of expected future values (in units of SDs of the distribution of the noise) as a function of recent data direction, arrow, global trend direction and block order.

Condition	Recent data direction					
	Recent-consistent		Recent-up		Recent-down	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Arrow absent						
Negative global trend						
Arrow absent-present	-0.508	0.720	0.587	1.102	-1.321	0.717
Arrow present-absent	-0.249	1.041	0.395	0.989	-0.692	1.058
Positive global trend						
Arrow absent-present	0.729	0.597	1.074	0.856	-0.513	1.191
Arrow present-absent	0.070	0.888	0.744	1.098	-0.442	0.901
Arrow present						
Negative global trend						
Arrow absent-present	-0.509	0.559	0.272	0.825	-1.001	1.041
Arrow present-absent	-0.297	1.102	0.379	1.017	-0.971	0.837
Positive global trend						
Arrow absent-present	0.243	0.923	0.632	0.996	-0.341	0.777
Arrow present-absent	0.320	0.821	0.913	0.815	-0.028	0.935

Note: Control-Arrow, $n = 16$; Arrow-Control, $n = 16$.

Table A7-2.

Study 8 means and standard deviations of the mean number of 'yes' responses to the outer probes as a function of arrow and block order.

Condition	<i>M</i>	<i>SD</i>
Arrow absent		
Arrow absent-present	0.407	0.111
Arrow present-absent	0.420	0.140
Arrow present		
Arrow absent-present	0.350	0.126
Arrow present-absent	0.417	0.130

Note: Control-Arrow, $n = 16$; Arrow-Control, $n = 16$.

Table A7-3.

Study 8 means and standard deviations of VAS scores for trials in which the probe was judged consistent with past data, as a function of probe location, arrow and block order.

Condition	Probe location			
	Central Probe		Outer Probes	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Arrow absent				
Arrow absent-present	74.026	18.081	59.420	12.210
Arrow present-absent	75.863	13.455	66.770	14.990
Arrow present				
Arrow absent-present	83.521	13.584	68.116	11.993
Arrow present-absent	78.538	11.131	64.858	12.724

Note: Control-Arrow, $n = 16$; Arrow-Control, $n = 16$.

Appendix 8

Data tables (means and standard deviations) for Study 9 analyses.

Table A8-1.

Study 9 means and standard deviations of mean location of expected future values (in units of SDs of the distribution of the noise) as a function of recent data direction, orientation, global trend direction and block order.

Condition	Recent data direction					
	Recent-consistent		Recent-up		Recent-down	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Horizontal						
Negative global trend						
Horizontal-Vertical	-0.106	0.856	0.419	0.773	-1.213	0.565
Vertical-Horizontal	-1.162	0.807	-0.472	0.995	-1.780	0.724
Positive global trend						
Horizontal-Vertical	0.171	0.811	1.181	0.843	-0.570	0.838
Vertical-Horizontal	0.620	0.697	1.502	0.589	0.176	0.735
Vertical						
Negative global trend						
Horizontal-Vertical	0.163	1.047	0.887	1.051	-0.863	0.653
Vertical-Horizontal	-0.742	0.736	-0.023	0.924	-1.676	0.610
Positive global trend						
Horizontal-Vertical	0.551	0.805	1.364	0.541	-0.091	0.583
Vertical-Horizontal	0.721	0.360	1.612	0.751	-0.211	1.067

Note: Horizontal-Vertical, $n = 20$; Vertical-Horizontal, $n = 20$.
Table A8-2.

Study 9 means and standard deviations of the mean number of ‘yes’ responses to the outer probes as a function of orientation and block order.

Condition	<i>M</i>	<i>SD</i>
Horizontal		
Horizontal-Vertical	0.401	0.084
Vertical-Horizontal	0.352	0.140
Vertical		
Horizontal-Vertical	0.422	0.107
Vertical-Horizontal	0.373	0.127

Note: Horizontal-Vertical, $n = 20$; Vertical-Horizontal, $n = 20$.

Table A8-3.

Study 9 means and standard deviations of VAS scores for trials in which the probe was judged consistent with past data, as a function of probe location, orientation and block order.

Condition	Probe location			
	Central Probe		Outer Probes	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Horizontal				
Horizontal-Vertical	71.259	12.750	58.214	9.292
Vertical-Horizontal	69.745	14.413	62.620	13.972
Vertical				
Horizontal-Vertical	69.841	14.358	58.649	11.885
Vertical-Horizontal	67.616	12.921	59.225	11.950

Note: Horizontal-Vertical, $n = 20$; Vertical-Horizontal, $n = 20$.

Appendix 9

Harold, Lorenzoni, Shipley & Coventry (2016)

Cognitive and psychological science insights to improve climate change data visualization. *Nature Climate Change*, 6, 1080-1089.

Cognitive and psychological science insights to improve climate change data visualisation

Author information

Jordan Harold*

School of Psychology and Tyndall Centre for Climate Change Research,
University of East Anglia, Norwich, NR4 7TJ, UK.

Jordan.Harold@uea.ac.uk

Irene Lorenzoni

School of Environmental Sciences, Tyndall Centre for Climate Change Research,
and Science, Society and Sustainability (3S) Research Group, University of East
Anglia, Norwich, NR4 7TJ, UK.

I.Lorenzoni@uea.ac.uk

Thomas F. Shipley

Department of Psychology, Temple University, Philadelphia PA 19122, USA.

tshipley@temple.edu

Kenny R. Coventry*

School of Psychology, University of East Anglia, Norwich, NR4 7TJ, UK.

k.coventry@uea.ac.uk

Corresponding author

Correspondence to: Jordan Harold or Kenny R. Coventry

Acknowledgements

This work was supported by a PhD Studentship from the School of Psychology, University of East Anglia (UEA) to J.H. and support from the Spatial Intelligence & Learning Centre (SILC), Temple University (SBE-1041707 from the National Science Foundation) including a travel grant to J.H. We would like to thank members of the Cognition Action Perception research group in the School of Psychology, UEA for their participation in a workshop to explore the scope of the review, and members of the Tyndall Centre for Climate Change Research, UEA for their feedback on how the presented guidelines could work in practice.

Contributions

J.H. and K.R.C. outlined the scope of the review with input from T.F.S. and I.L. The manuscript was drafted and prepared by J.H. with critical feedback from K.R.C., I.L. and T.F.S. All authors contributed to editing of the final manuscript.

Competing financial interests

The authors declare no competing financial interests.

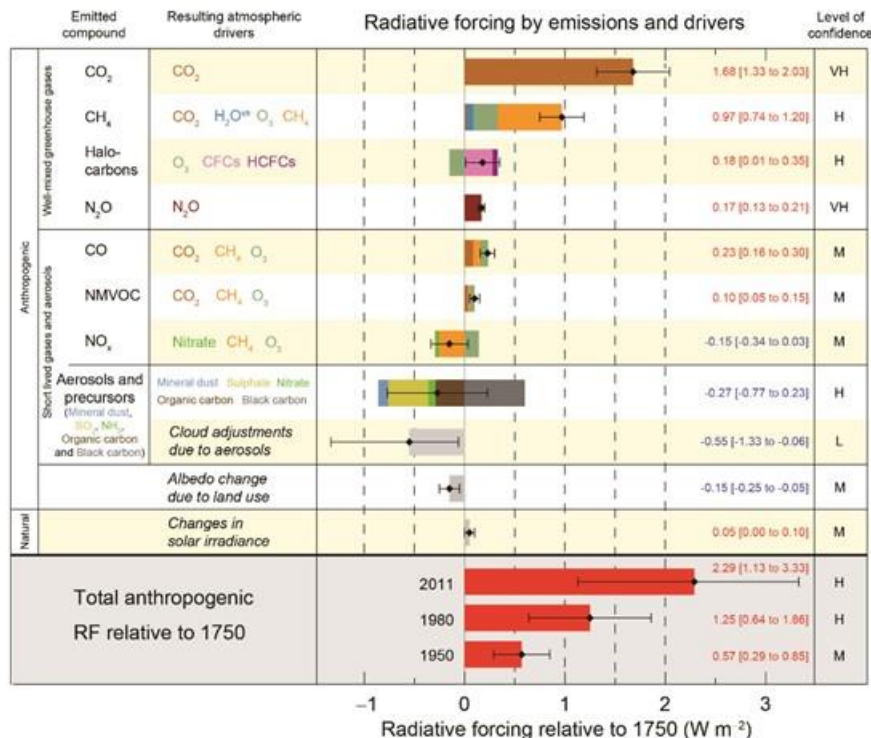
Abstract

Visualisation of climate data plays an integral role in the communication of climate change findings to both expert and non-expert audiences. The cognitive and psychological sciences can provide valuable insights into how to improve visualisation of climate data based on knowledge of how the human brain processes visual and linguistic information. We review four key research areas to demonstrate their potential to make data more accessible to diverse audiences: directing visual attention; visual complexity; making inferences from visuals; and the mapping between visuals and language. We present evidence-informed guidelines to help climate scientists increase the accessibility of graphics to non-experts, and illustrate how the guidelines can work in practice in the context of IPCC graphics.

Limiting the risks of severe impacts from climate change will require substantial changes in society to mitigate greenhouse gas emissions and adapt to a changing world¹. Scientific information is one factor among many that can influence decision-making to action change^{2,3} and there is an increasing demand for accessible and relevant climate data by decision-makers⁴. Global assessments of climate change by the Intergovernmental Panel on Climate Change (IPCC) provide important policy-relevant information. While summaries of these assessments are primarily aimed at experts working in government, they have been criticised for being inaccessible to non-experts, with particular focus on the complexity of language used in Summaries for Policy Makers (SPMs)^{5,6,7}. However, figures within SPMs (i.e. graphics of scientific information in the form of graphs, diagrams, thematic maps and other visuals), may also be inaccessible to non-experts (Fig. 1).

For example, viewers looking at graphics of climate model projections can confuse scenario uncertainty (i.e. unknown future societal choices) with model uncertainty⁸. There are challenges in visually synthesizing and representing uncertainty in climate knowledge, and diversity in normative judgements about the implications of such uncertainties⁹. Climate scientists may use different strategies to create meaning from climate science graphics than non-experts¹⁰. Furthermore, graphics of the same data represented in various styles have been shown to differentially influence judgements about future climate¹¹.

a.



b.

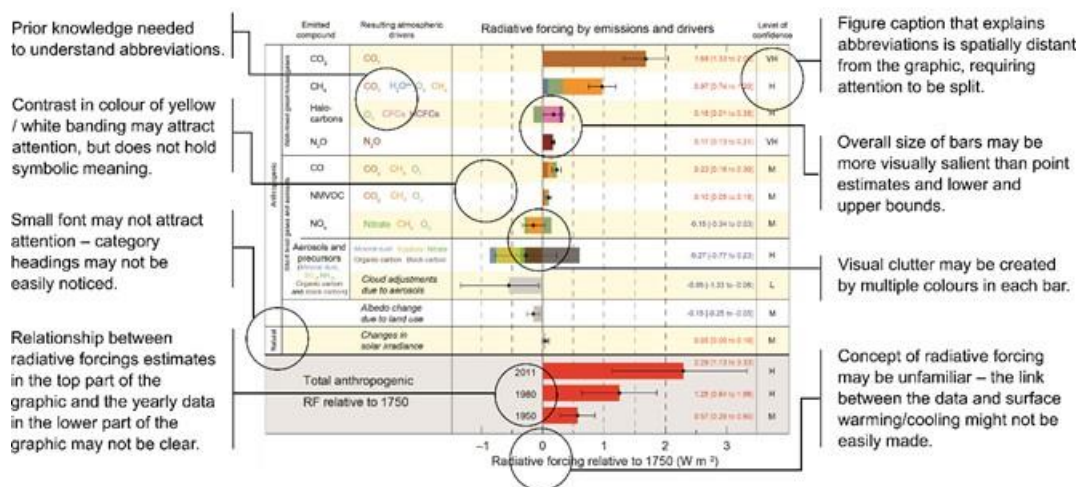


Figure 1. a. An example of a scientifically rigorous, policy-relevant IPCC graphic (caption below)⁹⁹. **b.** Aspects that might limit the accessibility of the graphic to non-expert audiences.

IPCC, AR5, Working Group 1, Figure SPM.5. Radiative forcing estimates in 2011 relative to 1750 and aggregated uncertainties for the main drivers of climate change. Values are global average radiative forcing (RF¹⁴), partitioned according to the emitted compounds or processes that result in a combination of drivers. The best estimates of the net radiative forcing are shown as black diamonds with corresponding uncertainty intervals; the numerical values are provided on the right of the figure, together with the confidence level in the net forcing (VH – very high, H – high, M – medium, L – low, VL – very low). Albedo forcing due to black

carbon on snow and ice is included in the black carbon aerosol bar. Small forcings due to contrails (0.05 W m^{-2} , including contrail induced cirrus), and HFCs, PFCs and SF_6 (total 0.03 W m^{-2}) are not shown. Concentration-based RFs for gases can be obtained by summing the like-coloured bars. Volcanic forcing is not included as its episodic nature makes it difficult to compare to other forcing mechanisms. Total anthropogenic radiative forcing is provided for three different years relative to 1750.

Visually representing climate data to inform decision-making can be challenging due to the multi-dimensionality of data, the diversity in users' needs across different stakeholder groups, and challenges and limitations in the use of software and tools to create graphics¹². However, graphics can, in principle, support thinking¹³ and support narratives when communicating with stakeholders¹⁴. Creating graphics of climate change data that overcome comprehension difficulties and avoid misconceptions has the potential to enhance climate change communications.

How can scientific graphics about climate change be made more accessible, while retaining their scientific integrity? This question has been posed by the IPCC as they look ahead to the Sixth IPCC Assessment Report¹⁵. In this review we consider research from the cognitive and psychological sciences to help answer this question. One of the goals of these disciplines is to understand how people comprehend written and visual information. We provide an overview of how people create meaning from graphical representations of data and highlight that intuitive design may not always correspond to best practice informed by evidence. We then consider four key areas: directing visual attention; reducing visual complexity; supporting inference-making; and integrating text with graphics. We present evidence-informed guidelines to support climate scientists in developing more accessible graphics, show how the guidelines can be applied in practice, and provide recommendations on how the IPCC might utilise these guidelines in the development of future reports.

We argue that improving accessibility to graphics of climate change data does not necessitate reducing or simplifying the *content* of the graphics per se (which might come with a risk of diluting the science), but can be achieved by supporting cognitive processing of the visual information.

Creating meaning from a scientific graphic

Graphics are often an effective way to communicate climate data - not only can they store and organise data efficiently, but they enable us to think about the data using visual perception¹³. Representing data visually can create patterns that the human visual system can easily process (e.g. the iconic ‘hockey-stick’ graph). However, graphics are not direct representations of reality; the meaning of the data they represent must be interpreted by the viewer. Therefore, prior to identifying how graphics of climate data might be made more accessible, we outline how the human brain creates meaning from a graphic.

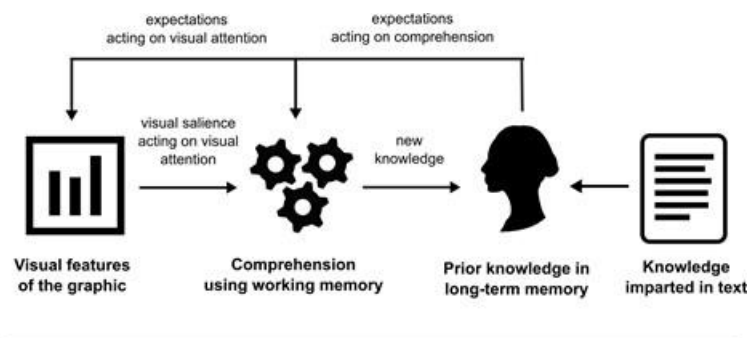
First, sensory processes direct the eyes to specific features of the graphic. Visual attention determines which features of the graphic the viewer looks at. Features that are visually salient (e.g. by virtue of their colour, shape, size) can draw the attention of the viewer – known as *bottom-up* visual processing. Conversely, the viewer’s expectations, driven by prior knowledge (their previous experience of the world, and their goal or reason for looking at the graphic), can also direct visual attention – *top-down* visual processing (Fig. 2a)¹⁶. As visual information is perceived from the features of the graphic, a mental representation of the information is created in memory. The nature of the mental representation is influenced by prior knowledge and goals and is constantly updated as the viewer visually explores the graphic¹³.

These cognitive processes are cyclical in nature; perceived and mentally represented information acts on expectations, which in turn direct further exploration of the graphic¹⁷. The human brain is thought to support cognition by constantly trying to match incoming sensory information against predictions of what to expect¹⁸. When perceived information matches our expectations, then comprehension is easy. Accessibility of a graphic can therefore be improved by matching visual features and prior knowledge (Fig. 2b).

a. Comprehension of graphics

Comprehension involves direction of visual attention, which can be driven by the visual features of the graphic and by the viewer's expectations from their prior knowledge.

Comprehension of attended information takes place in the context of prior knowledge using working memory, and enables inferences to be drawn from the data, creating new knowledge in long-term memory.



b. Improving accessibility

Graphics can be made more accessible and more easily understood by matching graphic parameters with parameters that influence or make up prior knowledge of the viewer.

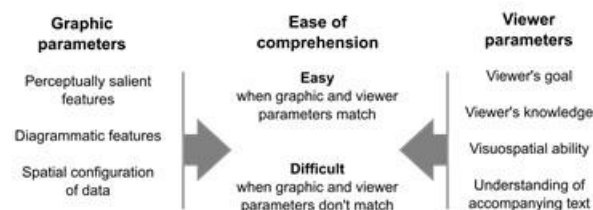


Figure 2: Conceptual overview of the process of graphic comprehension and approaches to improving accessibility.

Intuitive design \neq improved accessibility

Advances in computing and software technologies have enabled climate scientists to create a wide-range of visual representations of scientific data¹². In addition, such representations may offer the viewer flexibility in how the data are displayed via interaction with the graphic. Such advances offer the potential to better match graphic parameters to viewer parameters to improve accessibility. However, these advances also place demands on creators and viewers of graphics in terms of their competence in selecting effective visual representations of the data for the task at hand¹⁹.

Evidence suggests there may be limits to experts' self-awareness (metacognition) for creating or choosing effective visual representations of data. For example, some experts, as well as non-experts, show preferences for graphic features that can actually impair comprehension, such as realistic features²⁰, 3D features²¹ and extraneous variables in data²². Consequently, intuitions about good design practices may not always match best practice informed by cognitive principles, and viewer preferences may not always be predictive of ease of comprehension.

Conversely, designing graphics with cognitive principles in mind, and testing them with viewers, offers an empirical approach to improving the visual communication of climate science data.

Accessibility \neq loss of scientific rigour

The role of visual attention

To understand the details of a graphic we use our central vision, afforded by the fovea centralis, which provides greater acuity than our peripheral vision. The visual field of the fovea centralis is approximately two degrees of visual angle in diameter²³, meaning that when viewing an image from a distance of 60 cm (such as on a computer screen at about arm's length), our central vision covers an area approximately 2 cm wide. At any one moment in time our central vision can only focus on a limited area of a graphic. Therefore, we move our eye gaze to sample information from different spatial locations (Fig. 3a), and to build a detailed representation of the graphic as a whole we encode and retain information from these different spatial locations in memory.

Limited cognitive resources mean that only a fraction of the rich visual information entering the eyes at any given point in time is meaningfully processed and encoded to our internal representation in memory²⁴. Where to look, and what information to process, is directed by visual attention. Consequently, if important details in a graphic are not captured by our attention, they will not be processed by the brain and will not be drawn on to help comprehend and interpret the data in the graphic (Fig. 3b). Directing visual attention to important details can therefore make graphics more accessible by supporting viewers to look at aspects of the graphic that afford understanding.

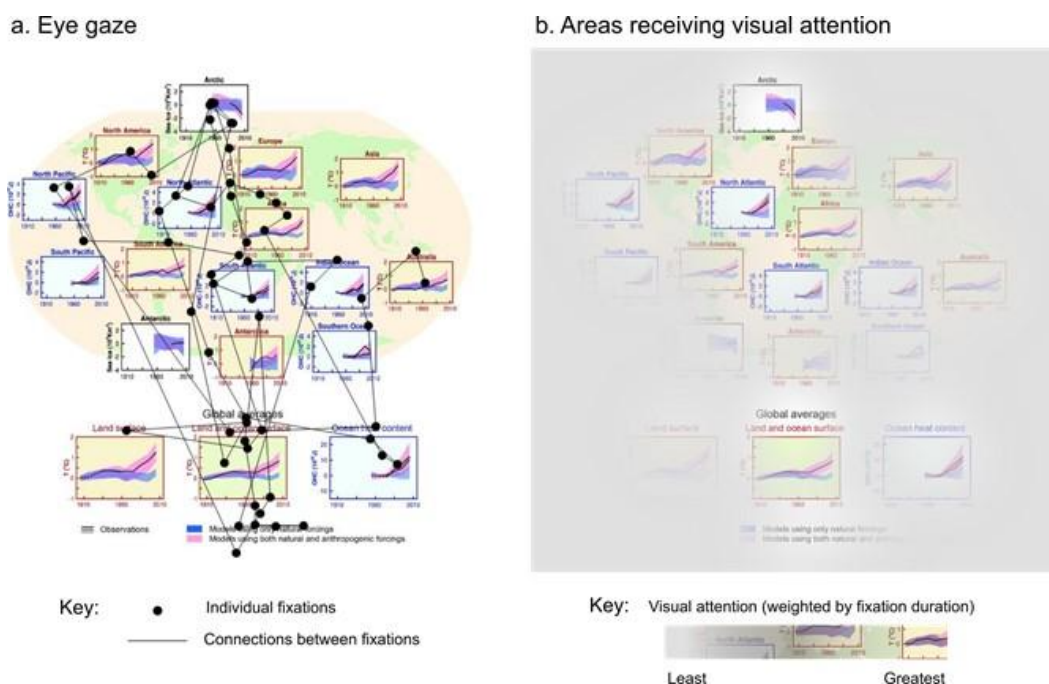


Figure 3. Example of visual attention for an IPCC figure for a non-expert viewer trying to interpret the graphic (measured using eye tracking: first 15 seconds of data shown). **a:** eye gaze shown as individual fixations and connections between fixations; **b:** areas receiving visual attention; computed from the locations of the fixations, weighted by the duration of each fixation. If visual features are not visually salient, they may not be attended to. In this example, the graphic’s legend receives little visual attention and some parts of the legend receive no visual attention at all.

Figure shown is IPCC, AR5, Working Group 1, Figure SPM.6.⁹⁹ Comparison of observed and simulated climate change based on three large-scale indicators in the atmosphere, the cryosphere and the ocean: change in continental land surface air temperatures (yellow panels), Arctic and Antarctic September sea ice extent (white panels), and upper ocean heat content in the major ocean basins (blue panels). Global average changes are also given. Anomalies are given relative to 1880–1919 for surface temperatures, 1960–1980 for ocean heat content and 1979–1999 for sea ice. All time-series are decadal averages, plotted at the centre of the decade. For temperature panels, observations are dashed lines if the spatial coverage of areas being examined is below 50%. For ocean heat content and sea ice panels the solid line is where the coverage of data is good and higher in quality, and the dashed line is where the data coverage is only adequate, and thus, uncertainty is larger. Model results shown are Coupled Model Intercomparison Project Phase 5 (CMIP5) multi-model ensemble ranges, with shaded bands indicating the 5 to 95% confidence intervals.

Directing attention by visual design

Visual properties that can capture attention by acting on bottom-up perceptual processing include colour, motion, orientation and size²⁵. In addition, there are well-documented ‘Gestalt’ principles governing how individual elements in a graphic are grouped together psychologically into meaningful entities²⁶. When elements of a graphic show a large degree of contrast in these properties, the contrasting visual information is automatically captured by attention and appears to ‘pop-out’ from the display (Fig. 4b-4d).

Another way to direct attention is through the use of arrows. Arrows are the symbolic visual equivalent of pointing gestures, which have a widely accepted meaning of ‘look here’ and are thought to direct attention automatically²⁷. They can therefore be particularly efficient visual cues to establish joint attention between the author and the viewer for specific features in a graphic (Fig. 4e). Of course arrows also have other uses – such as denoting motion or temporal change – and one has to be careful not to use arrows to denote different operations within the same graphic.

Using these properties in the visual design of climate science graphics can therefore help guide attention. Particular visual properties (or combinations of these properties) to direct attention may be more suited than others, depending on the context in which they are used.

Informed by human behaviour and neuroscience, computational models of ‘bottom-up’ visual attention have been able to accurately predict which features of an image are most likely to be attended to²⁸. Such models provide immediate assessments of visually salient features of a graphic, and might be useful to inform the design process²⁹. To check viewers’ actual visual attention for a graphic, eye-tracking can provide empirical evidence to inform visual design. For example, eye tracking has been used to observe differences in the eye movements of individuals who were successful or unsuccessful in solving a problem scenario depicted in a graphic; visual elements that supported problem solving could then be made more visually salient³⁰.

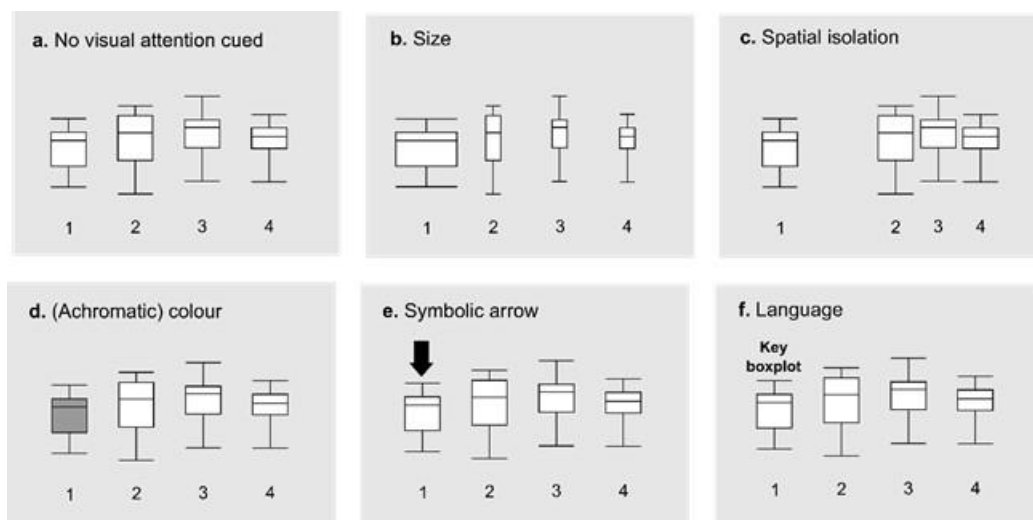


Figure 4. Schematic of properties known to direct visual attention that can be used in the design of graphics to help direct viewers' attention to important information.

Directing attention by informing expectation

The details that are looked at within a graphic can also be directed by expectations about the task at hand. For example, patterns of eye gaze are different when viewers search a graphic for a specific feature, compared to when they try to memorise the graphic as a whole³¹, or when a map is studied to learn routes as opposed to the overall layout³². Explicitly stating the intended task for which the graphic was created can help guide viewers' visual attention to appropriate information. Furthermore, prior knowledge about the data, and prior knowledge about the format or type of graphic chosen to represent the data, can also influence a viewer's cognition^{33,34}.

Research on the comprehension of meteorological charts has shown that providing viewers with relevant knowledge can support attention by directing it towards task-relevant features and away from task-irrelevant features³⁵.

Furthermore, making task-relevant features visually salient by adapting visual design may enhance performance once appropriate knowledge is provided³⁵.

Hence the interaction between bottom-up perceptual processing and top-down attentional control should be considered when designing graphics, with particular

consideration given to what knowledge the viewer needs to correctly interpret the data.

Handling complexity

Some climate science graphics are more visually complex than others. For example, ensemble datasets of climate models can be particularly complex and challenging to visualise³⁶. What is visual complexity, and how can complexity be handled to enable graphics to be more accessible? Possible components that might contribute towards defining and measuring visual complexity include the number of variables and/or data points in a graphic³⁷, the degree of uniformity of relationships represented by the data³³, or the degree to which the data are organised to make relevant relationships in the data easier to identify³⁸. However, while these components might be informative for simple graphics, they may not be easily applied across the diverse types of graphics used to communicate climate science, and may not always be predictive of comprehension. For example, in some instances an increasing number of data points might make patterns in the data more obvious.

An alternative proxy for visual complexity is ‘visual clutter’, where excess visual information, or a lack of organisation of that information, impairs cognition³⁹. Excess visual clutter can increase the time it takes to search for an item⁴⁰, increase errors in judgments⁴¹ and impair processing of language accompanying a graphic⁴². Computer models, based on principles of human cognition, can assess graphics for visual clutter and have been validated against viewers’ actual performance when undertaking simple tasks with graphics, such as searching for a specific feature³⁹. Although such models have yet to be established as offering diagnostic value in identifying comprehension problems with graphics, they can be useful to inform the design process by comparing different design options for a given graphic²⁹.

One approach to avoid unnecessary visual complexity is to only include information in a graphic that is absolutely needed for the intended purpose⁴³. However, climate science graphics may need to contain a certain level of detail or

information to maintain scientific integrity (i.e. to accurately represent the extent of, or limits to, scientific knowledge). Such graphics may still be visually complex in spite of only showing important information. While experts can integrate complex visual features into meaningful units of information (perceptual ‘chunks’), non-experts may lack such skills⁴⁴. Hence, segmenting information into chunks of appropriate size and difficulty, and guiding viewers’ attention to connections between these components could make comprehension of the data easier⁴⁵. However, such an approach should be taken with care. If the task expected of the viewer is to compare or contrast data represented in a graphic (known as ‘integrative tasks’), then this may be more easily performed when the data to be compared share representational similarities, such as close spatial proximity, or the same colour⁴⁶.

Supporting inference-making

Comprehension of a graphic of climate data goes beyond just perceptual processing of visual features. For example, enabling viewers to make relevant and scientifically robust inferences from data might be preferable to merely stating intended inferences in the accompanying text of a graphic. Furthermore, graphics are not only used to impart information, they can also be used to support sense-making and guide decision-making. In the context of the science-policy interface, this is indeed one of the goals of science communication and aligns with the IPCC’s remit of being policy-relevant and not policy prescriptive⁴⁷.

Improving accessibility to climate science graphics therefore involves supporting viewers to make appropriate inferences. Symbolic elements in diagrams, such as lines, boxes, crosses and circles can support inference-making about relationships in the data, based on their geometric properties⁴⁸. For example, lines indicate connections, while arrows can indicate dynamic, causal or functional information⁴⁹.

Inferences may also relate to the mappings between the visual features of the graphic and the data that they represent. Much of our cognition of conceptual ideas is thought to be metaphorical in nature⁵⁰. For example, *more* of something is

conceptualised in mind as *up*, and so temperature is said to be *rising*; similarly, financial concepts are used metaphorically in speech with regards to limiting carbon emissions, i.e. having a carbon *budget*. Using mappings that match natural or cultural metaphors can therefore aid cognition⁵⁰. For example, colour contains symbolic meaning, with red usually associated with ‘warm’ and blue with ‘cold’⁵¹, and indeed these colour choices are often used to represent temperature values in meteorological graphics. Metaphors often differ between cultures⁵² and so choice of metaphors should be informed by the target audience (see section below on tailoring graphics to different audiences).

How data are structured in a graphic can influence the type of information extracted, and in turn, what inferences are made about the data⁵³. For example, global climate projections are typically plotted as line graphs with time on the x-axis and the variable of interest (e.g. temperature anomaly) on the y-axis, which may direct viewers to consider given points in time and their associated temperature projections. Conversely, plotting temperature anomalies on the x-axis and time on the y-axis frames the data in terms of a projection of time for a given temperature threshold⁵⁴. Although in both cases the data are the same, the alternative graphical representations may result in viewers drawing different inferences.

Sometimes the viewer of a graphic may need to make inferences about the data that are not explicitly represented in the graphic. Examples include making inferences about the uncertainty of the data⁵⁵, relationships across multiple graphics⁵⁶, and relationships between a theory and data in a graphic⁵⁷. Such tasks involve spatial reasoning, i.e. the viewer must mentally infer information through spatial transformations⁵⁸. In such cases, inferences can be supported either by explicitly showing the inferences in the graphic (and so removing the need for spatial reasoning), or by supporting viewers’ spatial reasoning, for example by using text accompanying the graphic (see section below).

Using text to support cognition

Graphics of climate data are rarely used in isolation of accompanying text - text labels typically indicate the referents of the data, such as what the axes and data points represent. In accordance with norms of scientific reporting, captions provide contextual information and are placed under graphics, while the relevance of the graphic and inferences that can be drawn from it are placed in the body text, sometimes spatially distant from the graphic.

Separating text from graphics comes with a cognitive cost, known as the *spatial contiguity effect*⁵⁹. When there is distance between the spatial locations of the text and corresponding graphic, attention must be split between the two. The viewer must visually search for the corresponding elements (i.e. moving from text to graphic, or vice versa) and then integrate both sources of information. Viewers may not exert effort to do this and instead may simply treat text and graphics as independent units of information and read them independently of one another⁶⁰. However, when the distance between text and graphic is reduced, less searching is required, and connections can be more easily made, resulting in improved comprehension⁶¹. Tightly integrating text and graphic has been advocated as good design practice to support comprehension, i.e. embedding text *within* a graphic (Fig. 4f), or even embedding small graphics *within* text⁶².

Furthermore, language that accompanies a graphic has the potential not only to provide context, but also to influence thought about the spatial relationships of the properties of the graphic. Tasks involving spatial relationships might include comparisons of temperature anomalies at different spatial locations on a map, inferring trends in data from observed time-series data (which spatially plot x-y relationships), or comparing uncertainty ranges for future projections of climate under different scenarios. These tasks all involve spatial cognition, i.e. thinking about spatial relationships. Attending to linguistic information while looking at visual information is known to influence spatial cognition, such as supporting spatial reasoning⁶³. For example, a short sentence asking viewers to ignore extreme datapoints when looking at graphics of time series data results in participants attending to trends during encoding⁶⁴. Language can also influence

the extent to which a static visual is mentally animated and the manner in which it is animated⁶⁵, which again might help with spatial reasoning. Accompanying text can therefore support viewers in making appropriate spatial inferences from a graphic.

Tailoring graphics to different audiences

We have so far considered insights drawn from *general* principles of human cognition to help inform improved visual communication of climate science data. However, it is important to acknowledge that certain cognitive factors may differ between audience groups, and between individuals within those groups.

Colour is one area where there is marked individual and cultural variation. People who experience colour-blindness perceive colours differently from the general population and so colour choices for scientific graphics should be carefully chosen to avoid perceptual difficulties⁶⁶. The native language one speaks can also influence colour perception – the number of colour terms available in a language can influence colour discrimination⁶⁷, which might result in perceptual differences in the boundaries of colour-mapped data. Such problems can be avoided by using achromatic (e.g. greyscale) colour mappings in which data values are mapped to luminance rather than hue⁶⁸, or by using colour scales that enable easy differentiation of colour.⁶⁹

As well as perceptual differences, there are also group differences in higher-level cognitive skills, such as spatial reasoning. Experts often have strong spatial reasoning skills, as has been shown in the geosciences⁷⁰, whereas spatial reasoning by non-experts may depend on their general visuospatial abilities⁷¹. Moreover, how attention is directed across a page exhibits marked cultural variations, with reading direction in a language (e.g. English – left to right; Arabic – right to left) associated with the direction of attention in visuospatial tasks⁷².

Other differences are more tied to an individual's personal knowledge and experience. For example, prior experience can lead to a knowledge of 'where to look' and so can limit visual attention to specific spatial locations⁷³. Similarly, the

extent of prior knowledge about the data being visualised and prior experience using specific graphical formats can influence the ease with which inferences can be drawn from data⁷⁴. There can be trade-offs between using an unfamiliar graphical format that may be difficult to initially interpret but which efficiently represents a set of data, and a more familiar format whose structure can easily be grasped but which may provide an inefficient representation of the data³⁴.

Individuals may hold different and sometimes inaccurate mental models about complex scientific systems⁷⁵, such as the underlying physical principles of climate change⁷⁶. Understanding a viewer's existing mental model about the data and the systems from which the data originate can inform how they can best be supported to make scientifically robust inferences.

While comprehension of a graphic can be dependent on such factors outlined above, the underlying mechanisms responsible for human cognition are shared by everyone. Hence, general principles drawn from human cognition can inform approaches to improve the accessibility of graphics, but the specific way in which they are applied needs to be tailored. Consequently, testing of graphics is important to ensure they are comprehensible to achieve the desired communication goals^{8,13}.

Gaps in current knowledge

Despite advances in our understanding of the comprehension of graphics, there are important gaps in current knowledge that are of direct relevance to visualising climate data. Uncertainties of data can be difficult to communicate^{77,78}. Although general principles have been proposed for visually communicating probabilistic uncertainty, the deep uncertainties of climate change, in which knowledge and values are often disputed and outcomes are dependent on human behaviour, may not easily translate into visual representations⁷⁹. Further research is needed on how different visual representations of uncertainty might support or hinder decision-making⁸⁰ and the cognitive processes involved in such tasks.

To provide decision-makers with access to data tailored to their needs, researchers and climate service providers are exploring the use of interactive web-based

graphics, such as The Climate Explorer (part of the U.S. Climate Resilience Toolkit)⁸¹ and The IMPACT2C web-atlas⁸². Interaction, such as filtering or highlighting task-relevant information⁸³ has the potential to support comprehension. However, there can be large individual differences in the degree to which people use interactive functions and the extent to which they use these functions effectively⁸⁴; viewers require competence in meta-representational skills to make appropriate interactions¹⁹. Consequently, unless viewers have the required skills, there may be limits to how useful interactive graphics are to support comprehension and accessibility.

Both interactive graphics and animated graphics have been suggested to support the outreach of future IPCC assessments¹⁵. Research comparing static graphics with animated graphics is often confounded by additional information being provided in animated graphics; hence observed benefits of animation in some tasks may not be due to animation per se⁸⁵. In some cases animation may impair comprehension⁸⁶. Viewers may extract perceptually salient information rather than task-relevant information from animations^{87,88} and cognitive processing of the visual information may not be able to keep up with the pace of the animation^{87,89}. Animating graphics might be beneficial in specific situations if cognitive demands of processing the information are factored into the design of such graphics⁹⁰. Providing an element of user-control offers the potential to overcome some of these information processing limitations⁹¹. The decision to use an animated or interactive graphic over a static graphic should be informed by cognitive demands and task requirements, be designed taking cognitive principles into account, and be tested with viewers to check comprehension⁹².

Evidence-informed guidelines

Here we summarise the psychological insights considered by this review and provide associated guidelines that can help to improve accessibility of graphics of climate science (Table 1).

Table 1. Evidence-informed guidelines to improve accessibility of scientific graphics of climate science.

<i>Psychological insights</i>	<i>Associated guidelines to improve accessibility</i>
1. Intuitions about effective graphics do not always correspond to evidence-informed best practice for increasing accessibility ^{20,21,22}	Use cognitive and psychological principles to inform the design of graphics; test graphics during their development to understand viewers' comprehension of them ^{8,13}
Direct visual attention	
2. Visual attention is limited and selective – visual information in a graphic may or may not be looked at and/or processed by viewers ²⁴	Present only the visual information that is required for the communication goal at hand ⁴³ Direct viewers' visual attention to visual features of the graphic that support inferences about the data ⁹⁷
3. Salient visual features (where there is contrast in size, shape, colour or motion) can attract visual attention ^{25,26}	Make important visual features of the graphic perceptually salient so that they 'capture' the attention of the viewer ⁹⁷
4. Prior experience and knowledge can direct visual attention ^{34,35}	Choose and design graphics informed by viewers' familiarity and knowledge of using graphics

	<p>and their knowledge of the domain, i.e. knowledge about what the data represents⁴³</p> <p>Provide knowledge to viewers about which features of the graphic are important to look at, e.g. in text positioned close to the graphic (see Guideline 10)</p>
Reduce complexity	
<p>5. An excess of visual information can create visual clutter and impair comprehension^{40,41,42}</p>	<p>Only include information that is needed for the intended purpose of the graphic⁴³; break down the graphic into visual ‘chunks’, each of which should contain enough information for the intended task or message³⁸</p>
Support inference-making	
<p>6. Some inferences may require mental spatial transformations of the data⁵⁸; experts may have strong spatial reasoning skills⁷⁰, non-experts may not⁷¹</p>	<p>Remove or reduce the need for spatial reasoning skills by showing inferences directly in the graphic⁵⁶, and/or</p> <p>Support viewers in spatial reasoning, e.g. by providing</p>

	guidance in text ⁶⁴ (see Guideline 10)
7. The visual structure and layout of the data influences inferences drawn about the data ⁵³	Identify the most important relationships in the data that are to be communicated; consider different ways of structuring the data that enable the viewer to quickly identify these relationships ⁴³
8. Animating a graphic may help or hinder comprehension ^{85,86}	Decisions to create animated graphics should be informed by cognitive principles ⁹² ; consider providing user-control over the playback and speed of the animation ⁹¹
9. Conceptual thought often makes use of cultural metaphors ⁵⁰	Match the visual representation of data to metaphors that aid conceptual thinking, e.g. ‘up’ is associated with ‘good’ and ‘down’ is associated with ‘bad’; ⁵⁰ data with negative connotations may be easiest to understand if presented in a downwards direction ⁹⁸
Integrate text with graphics	

<p>10. When the graphic and the associated text are spatially distant, attention is split^{59,60}</p>	<p>Keep the graphic and accompanying text close together⁶², e.g. use text within a graphic and locate the graphic next to the accompanying body text</p>
<p>11. Language can influence thought about the graphic^{64,65}</p>	<p>Use text to help direct viewers' comprehension of the graphic, i.e. by providing key knowledge needed to interpret the graphic⁴³</p>

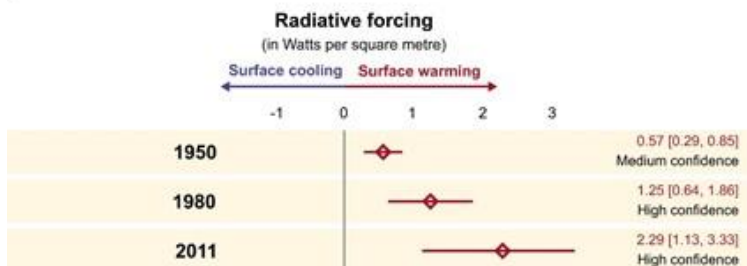
Guidelines in practice

To demonstrate how the guidelines can be applied in practice, we selected an IPCC SPM graphic (Fig. 1a) identified by IPCC authors (personal communication) as potentially challenging for comprehension. We first identified aspects that might hinder comprehension, especially when interpreted by non-experts (Fig. 1b). Drawing on the guidelines we then created a cognitively inspired version of the graphic, with the aim of making the data more widely accessible while retaining scientific integrity (Fig. 5 and Box 1).

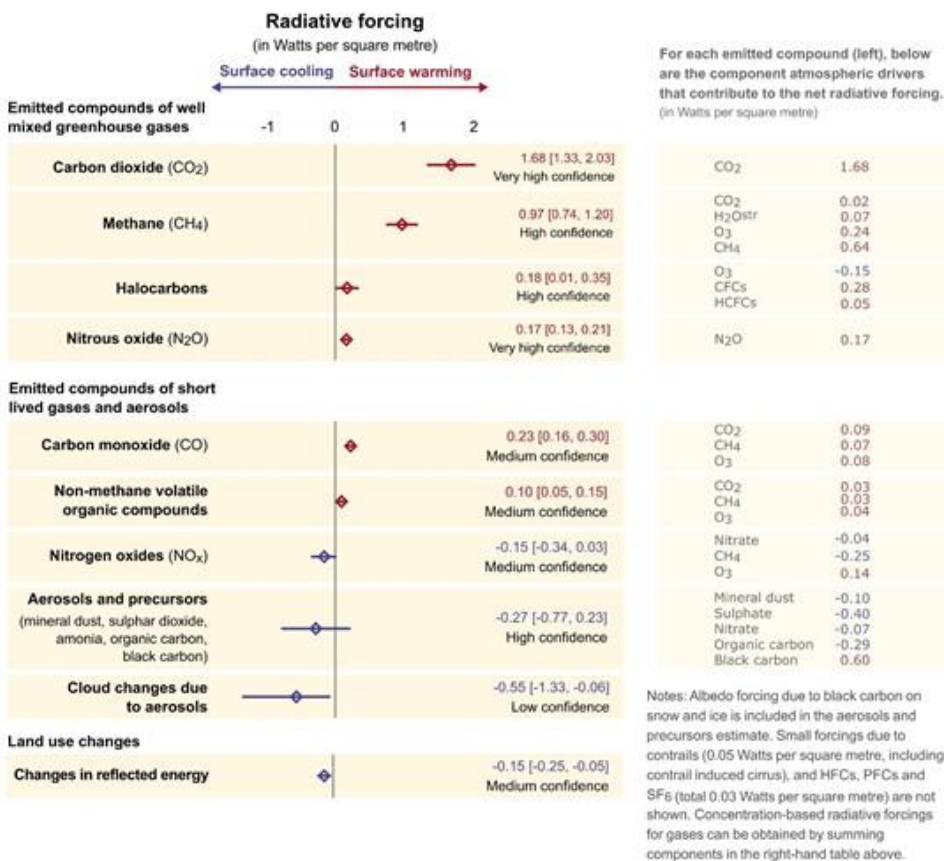
Key information

◆ Net radiative forcing best estimate and corresponding uncertainty interval 0.57 [0.29, 0.85] Medium confidence Plotted values and qualitative degree of confidence in the net radiative forcing (based on the type, amount, quality, and consistency of evidence)

Human activities: total radiative forcing for 1950, 1980 and 2011 (all relative to 1750)



Contributions to the total radiative forcing caused by human activities for 2011 (relative to 1750)



Natural causes: radiative forcing for 2011 (relative to 1750)

◆ Changes in solar irradiance 0.05 [0.00, 0.10] Medium confidence Notes: Volcanic forcing is not shown as its episodic nature makes it difficult to compare to other forcing mechanisms.

Figure 5. | A cognitively inspired version of IPCC AR5 WG1 SPM Figure SPM.6⁹⁹, using the guidelines in Table 1 to increase accessibility while maintaining scientific rigour (see also Box 1).

Box 1 | Guidelines used in the cognitively inspired version of IPCC AR5 WG1 SPM Figure SPM.6.

The cognitively inspired version provides knowledge of the meaning of all abbreviations (guideline 11); breaks down information into ‘chunks’ to reduce complexity and clutter (guideline 5); uses larger font size for headings, relative to other text, to attract attention (guideline 2 and 3); uses contrast in colour to encourage attention of the distinction between human and natural radiative forcings (guideline 3); shows the relationship between the 2011 total and the contributions to the total (guideline 7); integrates the caption text within the graphic to reduce the need for splitting attention (guideline 10); plots only point estimates and uncertainty ranges, i.e. removes bars, to reduce clutter and encourage thinking about the best estimate and uncertainty (guidelines 3 and 5); removes the need for multiple colours to represent each compound to reduce clutter (guideline 5); and uses text, and colour as a metaphor, to support understanding of link between the data and surface warming/cooling (guidelines 4,9,11).

We tested the alternative version of the graphic (Fig. 5) and the original (Fig. 1a) on a sample of experts (ten climate change researchers) and non-experts (ten psychology researchers). Eighty percent of participants indicated a preference for the cognitively inspired version, significantly more than expected by chance against the null hypothesis of there being no difference in preferences, exact binomial $p = .012$ (two-tailed). Such user-testing can help inform the development of graphics as part of an iterative design cycle.

Creating accessible graphics

There is the potential to develop improved scientific graphics of climate change data that are cognitively-inspired and easier to comprehend. This goal in particular aligns with the IPCC’s desire to make outputs of future reports more accessible and user-friendly to diverse audiences⁹³.

In addition, the ease of accessibility of graphics of climate science also has implications for how society might make best use of scientific knowledge. There have been calls for climate scientists to take participatory roles in co-productive frameworks alongside stakeholders to help inform societal decision-making⁹⁴.

Graphics of climate data that are accessible to all parties involved could support improved engagement, dialogue and decision-making between scientists, policy-makers, practitioners, communities and publics. Climate service providers (who supply tailored climate knowledge to decision-makers) often use graphics to communicate findings, and although the communication goals and intended audience may be much more specific in these contexts than the global assessments made by the IPCC, data visualisation challenges remain⁹⁵.

While the science underpinning graphic comprehension is still developing, the guidelines presented in this review provide a useful reference for climate scientists to apply psychological and cognitive insights when creating graphics of data. However, as individuals and groups can differ, there is no substitute for empirically testing graphics with the target audience. Such testing need not be costly or time-consuming. Asking people to look at and interpret drafts of graphics can indicate if graphics are broadly understandable or not. Furthermore, rich diagnostic evidence afforded by eye tracking can indicate the efficiency of comprehension and can identify reasons why comprehension is impaired, such as assessing whether task-relevant information is visually salient or not. Informed by such evidence, appropriate adjustments to graphics can be made and then they can be re-tested.

Greater collaboration between the climate change research community, the psychology and cognitive science community and those working in associated disciplines, could help to realise such an approach. For example, as the IPCC looks ahead to their Sixth Assessment Report, there is an opportunity for the IPCC to open up the review process and ask these communities for feedback on drafts of SPM graphics. Climate scientists and psychologists could also jointly develop cognitively-inspired graphics of climate data, which are both accessible and scientifically robust, for use in outputs outside of the formal IPCC process (so-called 'derivative products'). Similar collaborations between research communities have led to improved communication in related fields such as cartography⁹⁶ and geoscience⁷⁰.

Graphics of climate data are integral to scientific assessments of climate change, but only support communication and decision-making *if* they are understood. Empirically testing graphics and applying insights from the science of human cognition to help overcome comprehension problems, offers the potential to make climate science knowledge more accessible to decision-makers in society, while also retaining the integrity of the scientific data and evidence on which they are based.

References

- 1 IPCC *Climate Change 2014: Synthesis Report* (eds Pachauri, R. K. *et al.*) (Cambridge Univ. Press, 2014).
- 2 Clayton, S. *et al.* Psychological research and global climate change. *Nature Clim. Change* **5**, 640-646 (2015).
- 3 Dilling, L. & Lemos, M. C. Creating usable science: Opportunities and constraints for climate knowledge use and their implications for science policy. *Glob. Environ. Change* **21**, 680-689 (2011).
- 4 Overpeck, J. T., Meehl, G. A., Bony, S. & Easterling D. R. Climate data challenges in the 21st century. *Science* **331**, 700-702 (2011).
- 5 Barkemeyer, R., Dessai, S., Monge-Sanz, B., Renzi, B. G. & Napolitano, G. Linguistic analysis of IPCC summaries for policymakers and associated coverage. *Nature Clim. Change* **6**, 311-317 (2016).
- 6 Hollin, G. J. S. & Pearce, W. Tension between scientific certainty and meaning complicates communication of IPCC reports. *Nature Clim. Change* **5**, 753-756 (2015).
- 7 Budescu, D. V., Por, H. H., Broomell, S. B. & Smithson, M. The interpretation of IPCC probabilistic statements around the world. *Nature Clim. Change* **4**, 508-512 (2014).

8 McMahon, R., Stauffacher, M. & Knutti, R. The unseen uncertainties in climate change: reviewing comprehension of an IPCC scenario graph. *Clim. Change* **133**, 141-154 (2015).

9 Mahony, M. & Hulme, M. The colour of risk: An exploration of the IPCC's "burning embers" diagram. *Spontaneous Generations: Journal for the History and Philosophy of Science* **6**, 75-89 (2012).

10 Stofer, K. & Che, X. Comparing experts and novices on scaffolded data visualizations using eye-tracking. *J. Eye Mov. Res.* **7**(5), 2 (2014).

11 Daron, J. D., Lorenz, S., Wolski, P., Blamey, R. C. & Jack, C. Interpreting climate data visualisations to inform adaptation decisions. *Clim. Risk Manage.* **10**, 17-26 (2015).

12 Nocke, T., Sterzel, T., Böttinger, M. & Wrobel, M. (2008) in *Digital Earth Summit on Geoinformatics 2008: Tools for Global Change Research*. (eds Ehlers M. et al.) 226-232 (Wichmann, 2008).

13 Hegarty, M. The cognitive science of visual-spatial displays: Implications for design. *Top. Cogn. Sci.* **3**, 446-474 (2011).

14 O'Neill, S., Williams, H. T. P., Kurz, T., Wiersma, B. & Boykoff, M. Dominant frames in legacy and social media coverage of the IPCC Fifth Assessment Report. *Nature Clim. Change* **5**, 380-385 (2015).

15 *IPCC Expert Meeting on Communication Meeting Report* (eds Lynn, J. et al.) (World Meteorological Organization, 2016).

16 Pinker, S. in *Artificial Intelligence and the Future of Testing* (ed Freedle, R.) 73-126 (Lawrence Erlbaum Associates, 1990).

17 Neisser, U. *Cognition and Reality: Principles and Implications of Cognitive Psychology*. (W. H. Freeman & Company, 1976).

18 Clark, A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* **36**, 181-204 (2013).

- 19 diSessa, A. A. Metarepresentation: Native competence and targets for instruction. *Cognition Instruct.* **22**, 293-331 (2004).
- 20 Smallman, H. S. & St. John, M. Naïve realism: Misplaced faith in realistic displays. *Ergon. Des.* **13**, 6-13 (2005).
- 21 Zacks, J., Levy, E., Tversky, B. & Schiano, D. J. Reading bar graphs: Effects of extraneous depth cues and graphical context. *J. Exp. Psychol. Appl.* **4**, 119-138 (1998).
- 22 Hegarty, M., Smallman, H. S., Stull, A. T. & Canham, M. S. Naïve cartography: How intuitions about display configuration can hurt performance. *Cartographica* **44**, 171-186 (2009).
- 23 Rayner, K. Eye movements and attention in reading, scene perception, and visual search. *Q. J. Exp. Psychol.* **62**, 1457-1506 (2009).
- 24 Simons, D. J. & Chabris, C. F. Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception* **28**, 1059-1074 (1999).
- 25 Wolfe, J. M. & Horowitz, T. S. What attributes guide the deployment of visual attention and how do they do it? *Nature Rev. Neurosci.* **5**, 495-501 (2004).
- 26 Bruce, V., Green, P. R., & Georgeson, M. A. *Visual Perception: Physiology, Psychology, & Ecology*. (Psychology Press, 2003).
- 27 Hommel, B., Pratt, J., Colzato, L. & Godijn, R. Symbolic control of visual attention. *Psychol. Sci.* **12**, 360-365 (2001).
- 28 Itti, L. & Koch, C. Computational modelling of visual attention. *Nature Rev. Neurosci.* **2**, 194-203 (2001).
- 29 Rosenholtz, R., Dorai, A. & Freeman, R. Do predictions of visual perception aid design? *ACM Trans. Appl. Percep. (TAP)* **8**, 12 (2011).
- 30 Grant, E. R. & Spivey, M. J. Eye movements and problem solving: Guiding attention guides thought. *Psychol. Sci.* **14**, 462-466 (2003).

- 31 Henderson, J. M., Weeks, P. A. & Hollingworth, A. The effects of semantic consistency on eye movements during complex scene viewing. *J. Exp. Psychol. Hum. Percept. Perform.* **25**, 210–228 (1999).
- 32 Brunyé, T. T. & Taylor, H. A. When goals constrain: Eye movements and memory for goal-oriented map study. *Appl. Cognitive Psych.* **23**, 772-787 (2009).
- 33 Carpenter, P. A. & Shah, P. A model of the perceptual and conceptual processes in graph comprehension. *J. Exp. Psychol. Appl.* **4**, 75-100 (1998).
- 34 Peebles, D. & Cheng, P. C. H. Modeling the effect of task and graphical representation on response latency in a graph reading task. *Hum. Factors* **45**, 28-46 (2003).
- 35 Hegarty, M., Canham, M. S. & Fabrikant, S. I. Thinking about the weather: How display salience and knowledge affect performance in a graphic inference task. *J. Exp. Psychol. Learn. Mem. Cogn.* **36**, 37-53 (2010).
- 36 Potter, K. *et al.* Ensemble-vis: A framework for the statistical visualization of ensemble data. *IEEE International Conference on Data Mining Workshops*, 233-240 (2009).
- 37 Meyer, J., Shinar, D. & Leiser, D. Multiple factors that determine performance with tables and graphs. *Hum. Factors* **39**, 268-286 (1997).
- 38 Shah P., Mayer, R. E. & Hegarty, M. Graphs as aids to knowledge construction: Signaling techniques for guiding the process of graph comprehension. *J. Educ. Psychol.* **91**, 690-702 (1999).
- 39 Rosenholtz, R., Li, Y. & Nakano, L. Measuring visual clutter. *J. Vision* **7**, 17 (2007).
- 40 Neider, M. B. & Zelinsky, G. J. Cutting through the clutter: Searching for targets in evolving complex scenes. *J. Vision* **11**, 7 (2011).
- 41 Baldassi, S., Megna, N. & Burr, D. C. Visual clutter causes high-magnitude errors. *PLoS Biol.* **4**, e56 (2006).

- 42 Coco, M. I. & Keller, F. in *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (eds Taatgen, N. & van Rijn, H.) 274-279 (Cognitive Science Society, 2009).
- 43 Kosslyn, S. M. *Graph design for the eye and mind*. (OUP, 2006).
- 44 Chase, W. G. & Simon, H. A. Perception in chess. *Cognitive Psychol.* **4**, 55-81 (1973).
- 45 Gobet, F. Chunking models of expertise: Implications for education. *Appl. Cognitive Psych.* **19**, 183-204 (2005).
- 46 Wickens, C. D. & Carswell, C. M. The proximity compatibility principle: Its psychological foundation and relevance to display design. *Hum. Factors* **37**, 473-494 (1995).
- 47 *Decisions Taken with Respect to the Review of IPCC Processes and Procedures: Communications Strategy* (IPCC, 2012);
http://www.ipcc.ch/meetings/session35/IAC_CommunicationStrategy.pdf
- 48 Tversky, B. in *Handbook of Higher-Level Visuospatial Thinking* (eds Shah, P. & Miyake, A.) 1-34 (Cambridge Univ. Press, 2005).
- 49 Heiser, J. & Tversky, B. Arrows in comprehending and producing mechanical diagrams. *Cognitive Sci.* **30**, 581-592 (2006).
- 50 Lakoff, G. & Johnson, M. The metaphorical structure of the human conceptual system. *Cognitive Sci.* **4**, 195-208 (1980).
- 51 Ho, H. N., Van Doorn, G. H., Kawabe, T., Watanabe, J. & Spence, C. Colour-temperature correspondences: When reactions to thermal stimuli are influenced by colour. *PloS One* **9**, e91854 (2014).
- 52 Kövecses, Z. *Metaphor in culture: Universality and variation*. (Cambridge University Press, 2005).
- 53 Shah, P. & Carpenter, P. A. Conceptual limitations in comprehending line graphs. *J. Exp. Psychol. Gen.* **124**, 43-61 (1995).

- 54 Joshi, M., Hawkins, E., Sutton R., Lowe, J. & Frame, D. Projections of when temperature change will exceed 2 °C above pre-industrial levels. *Nature Clim. Change* **1**, 407-412 (2011).
- 55 Trickett, S. B., Trafton, J. G., Saner, L. & Schunn, C. D. in *Thinking with Data* (eds Lovett, M. C. & Shah, P.) 65-85 (Psychology Press, 2007).
- 56 Trafton, J. G. *et al.* Turning pictures into numbers: Extracting and generating information from complex visualizations. *Int. J. Hum.-Comp. St.* **53**, 827-850 (2000).
- 57 Trafton, J. G., Trickett, S. B. & Mintz, F. E. Connecting internal and external representations: Spatial transformations of scientific visualizations. *Found. Sci.* **10**, 89-106 (2005).
- 58 Trafton, J. G., Marshall, S., Mintz, F. & Trickett, S. B. in *Diagrammatic Representation and Inference, Volume 2317 Lecture Notes in Computer Science* (eds Hegarty, M., Meyer, B. & Narayanan, N. H.) 206-220. (Springer Berlin, 2002).
- 59 Mayer, R.E. *Multimedia Learning* (Cambridge Univ. Press, 2009).
- 60 Holsanova, J., Holmberg, N. & Holmqvist, K. Reading information graphics: The role of spatial contiguity and dual attentional guidance. *Appl. Cognitive Psychol.* **23**, 1215-1226 (2009).
- 61 Ginns, P. Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learn. Instr.* **16**, 511-525 (2006).
- 62 Tufte, E. R. *Beautiful Evidence* (Graphics Press, 2006).
- 63 Loewenstein, J. & Gentner, D. Relational language and the development of relational mapping. *Cognitive Psychol.* **50**, 315-353 (2005).
- 64 Harold, J., Coventry, K. R., Lorenzoni, I. & Shipley, T. F. in *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (eds Noelle, D. C. *et al.*) 872-877. (Cognitive Science Society, 2015).

- 65 Coventry, K. R., Christophel, T. B., Fehr, T., Valdés-Conroy, B. & Herrmann, M. Multiple routes to mental animation: Language and functional relations drive motion processing for static Images. *Psychol. Sci.* **24**, 1379-1388 (2013).
- 66 Light, A. & Bartlein, P. J. The end of the rainbow? Color schemes for improved data graphics. *Eos* **85**, 385-391 (2004).
- 67 Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B. & Kuipers, J. R. Unconscious effects of language-specific terminology on preattentive color perception. *PNAS* **106**, 4567-4570 (2009).
- 68 Moreland, K. in *Proceedings of the 5th International Symposium on Advances in Visual Computing: Part II* (eds Bebis, G. et al.) 92-103 (Springer-Verlag Berlin, 2009).
- 69 Harrower, M. & Brewer, C. A. ColorBrewer.org: An online tool for selecting colour schemes for maps. *Cartogr. J.* **40**, 27-37 (2003).
- 70 Shipley, T. F., Tikoff, B., Ormand, C. & Manduca, C. Structural geology practice and learning, from the perspective of cognitive science. *J. Struct. Geol.* **54**, 72-84 (2013).
- 71 Hambrick, D. Z. et al. A test of the circumvention-of-limits hypothesis in scientific problem solving: The case of geological bedrock mapping. *J. Exp. Psychol. Gen.* **141**, 397-403 (2012).
- 72 Shaki S., Fischer M. H. & Petrusic W. M. Reading habits for both words and numbers contribute to the SNARC effect. *Psychon. B. Rev.* **16**, 328-331 (2009).
- 73 Torralba, A., Oliva, A., Castelano, M. S. & Henderson, J. M. Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychol. Rev.* **113**, 766-786 (2006).
- 74 Ratwani, R. M. & Trafton, J. G. (2008). Shedding light on the graph schema: Perceptual features versus invariant structure. *Psychon. B. Rev.* **15**, 757-762 (2008).
- 75 Gentner, D. & Gentner, D. R. in *Mental Models* (eds Gentner, D. & Stevens, A. L.) 99-129 (Lawrence Erlbaum Associates, 1983).

- 76 Sterman, J. D. & Sweeney, L. B. Understanding public complacency about climate change: Adults' mental models of climate change violate conservation of matter. *Clim. Change* **80**, 213-238 (2007).
- 77 Gigerenzer, G., Hertwig, R., van den Broek, E., Fiasolo, B. & Katsikopoulos, K. V. "A 30% chance of rain tomorrow": How does the public understand probabilistic weather forecasts? *Risk Anal.* **25**, 623-629 (2005).
- 78 Budescu, D. V., Broomell, S. & Por, H. H. Improving communication of uncertainty in the reports of the Intergovernmental Panel on Climate Change. *Psychol. Sci.* **20**, 299-308 (2009).
- 79 Spiegelhalter, D., Pearson, M. & Short, I. Visualizing uncertainty about the future. *Science* **333**, 1393-1400 (2011).
- 80 Andrienko, G. *et al.* Space, time and visual analytics. *Int. J. Geogr. Inf. Sci.* **24**, 1577-1600 (2010).
- 81 U.S. Climate Resilience Toolkit, The Climate Explorer.
<https://toolkit.climate.gov/climate-explorer2/>
- 82 IMPACT2C web-atlas. <https://www.atlas.impact2c.eu/>
- 83 Crampton, J. W. Interactivity types in geographic visualization. *Cartogr. Geogr. Inf. Sci.* **29**, 85-98 (2002).
- 84 Cohen, C. A. & Hegarty, M. Individual differences in use of external visualisations to perform an internal visualisation task. *Appl. Cognitive Psych.* **21**, 701-711 (2007).
- 85 Tversky, B., Morrison, J. B. & Betrancourt, M. Animation: can it facilitate? *Int. J. Hum.-Comput. St.* **57**, 247-262 (2002).
- 86 Mayer, R. E., Hegarty, M., Mayer, S. & Campbell, J. When static media promote active learning: Annotated illustrations versus narrated animations in multimedia instruction. *J. Exp. Psychol. Appl.* **11**, 256-265 (2005).
- 87 Lowe, R. K. Extracting information from an animation during complex visual learning. *Eur. J. Psychol. Educ.* **14**, 225-244 (1999).

- 88 Lowe, R. K. Animation and learning: Selective processing of information in dynamic graphics. *Learn. Instr.* **13**, 157-176 (2003).
- 89 Hegarty, M., Kriz, S. & Cate, C. The roles of mental animations and external animations in understanding mechanical systems. *Cognition Instruct.* **21**, 325-360 (2003).
- 90 Griffin, A. L., MacEachren, A. M., Hardisty, F., Steiner, E. & Li, B. A comparison of animated maps with static small-multiple maps for visually identifying space-time clusters. *Ann. Assoc. Am. Geogr.* **96**, 740-753 (2006).
- 91 Betrancourt, M. in *The Cambridge Handbook of Multimedia Learning* (ed Mayer, R. E.) 287-296 (Cambridge Univ. Press, 2005).
- 92 Shipley, T. F., Fabrikant, S. I. & Lautenschütz, A. K. in *Cognitive and Linguistic Aspects of Geographic Space* (eds Raubal, M., Mark, D. M. & Frank, A. U.) 259-270 (Springer-Verlag Berlin, 2013).
- 93 *Report of the 41st Session of the IPCC.* (IPCC, 2015);
http://ipcc.ch/meetings/session41/final_report_p41.pdf
- 94 Rapley, C. G., et al. *Time for Change? Climate Science Reconsidered, Report of the UCL Policy Commission on Communicating Climate Science* (2014).
- 95 Davis, M., Lowe, R., Steffen, S., Doblas-Reyes, F. & Rodó, X. in *Communicating Climate-Change and Natural Hazard Risk and Cultivating Resilience* (eds Drake, J. L., Kontar, Y. Y., Eichelberger, J. C., Rupp, T. S. & Taylor, K. M.) 95-113 (Springer International Publishing, 2016).
- 96 Fabrikant, S. I., Hespanha, S. R. & Hegarty, M. Cognitively inspired and perceptually salient graphic displays for efficient spatial inference making. *Ann. Assoc. Am. Geogr.* **100**, 13-29 (2010).
- 97 Kosslyn, S. M. Understanding charts and graphs. *Appl. Cognitive Psych.* **3**, 185-225 (1989).
- 98 Meier, B. P. & Robinson, M. D. Why the sunny side is up: Associations between affect and vertical position. *Psychol. Sci.* **15**, 243-247 (2004).

99 IPCC *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. *et al.*) (Cambridge Univ. Press, 2013).

Appendix 10

Harold, Coventry, Lorenzoni, & Shipley (2015)

Harold, J., Coventry, K. R., Lorenzoni, I., & Shipley, T. F. Making Sense of Time-Series Data: How Language Can Help Identify Long-Term Trends. In Noelle, D. C., Dale, R., Warlaumont, A. S., Yoshimi, J., Matlock, T., Jennings, C. D., & Maglio, P. P. (Eds.) (2015). *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*. Austin, TX: Cognitive Science Society; pp 872-77.

Making Sense of Time-Series Data: How Language Can Help Identify Long-Term Trends

Jordan Harold (jordan.harold@uea.ac.uk)

School of Psychology and Tyndall Centre for Climate Change Research,
University of East Anglia, Norwich, UK

Kenny R. Coventry (k.coventry@uea.ac.uk)

School of Psychology, University of East Anglia, Norwich, UK

Irene Lorenzoni (i.lorenzoni@uea.ac.uk)

School of Environmental Sciences and Tyndall Centre for Climate Change
Research, University of East Anglia, Norwich, UK

Thomas F. Shipley (tshipley@temple.edu)

Department of Psychology, Temple University, Philadelphia, USA

Abstract

Real-world time-series data can show substantial short-term variability as well as underlying long-term trends. Verbal descriptions from a pilot study, in which participants interpreted a real-world line graph about climate change, revealed that trend interpretation might be problematic (Experiment 1). The effect of providing a graph interpretation strategy, via a linguistic warning, on the encoding of long-term trends was then tested using eye tracking (Experiment 2). The linguistic warning was found to direct visual attention to task-relevant information thus enabling more detailed internal representations of the data to be formed. Language may therefore be an effective tool to support users in making appropriate spatial inferences about data.

Keywords: graph comprehension; language; visual attention

Line graphs can be a powerful communication tool to visually demonstrate important relationships in time-series data. They are ubiquitous in everyday life and graph interpretation is considered an important skill for a scientifically literate society (Glazer, 2011). Many types of real-world data exhibit substantial short-term variability as well as long-term trends, e.g. global mean surface temperature records (IPCC, 2013), share prices (Schwert, 2011), and incidence of certain diseases (e.g. Subak, 2003). In visualizations of such data, can users efficiently and accurately identify underlying long-term trends? If not, how might users be supported in doing so?

Comprehension of graphs involves an interaction between bottom-up sensory processes and top-down cognitive constraints, and is thought to involve two key cyclical processes (Carpenter & Shah, 1998; Freedman & Shah, 2002). First, users construct an internal representation of the display by encoding perceptual features of the graph, guided by prior knowledge. Then knowledge is applied to integrate the representation into a coherent mental model. If relevant information is represented directly in the graph and can be easily linked with existing knowledge, this integration phase is comparatively effortless. However, if information is not explicitly represented in the graph and/or the user lacks the

required knowledge to form an accurate model, or cannot easily access the required knowledge, then comprehension is likely to require much more effort.

For example, a climate scientist will know to consider the long-term trend when interpreting temperature records and so may effortlessly transform and encode visual features from the data that support a representation of the long-term trend. In contrast, a climate science ‘novice’ may encode visual features that are explicitly represented in the graph, such as the amplitude of peaks or troughs, which may support an understanding of short-term fluctuations, but make inferences about the long-term trend rather effortful and less likely. Hence, graphs that organize and structure data, such that emergent visual properties explicitly reveal important relationships, e.g. based on Gestalt laws, may be particularly effective (Kosslyn, 1989; Zacks & Tversky, 1999), by reducing the cognitive effort that might otherwise be needed (Hegarty, 2011).

Although a line graph may be a single unit by the Gestalt law of connectedness (Ali & Peebles, 2013), a complex line may be decomposed into parts or ‘chunks’, based on local curvature extrema (Hoffman & Richards, 1984). Time-series datasets that show significant short-term variability may have numerous curvature extrema (e.g. trend reversals) creating multiple visual chunks. These chunks may serve as units on which inferential processes, required for interpretation, act (Freedman & Shah, 2002).

Trend reversals can increase study time, and also increase local content and decrease global content of verbal and written interpretations of line graphs (Carswell, Emery, & Lonon, 1993). In this study it was hypothesized that each set of continuous non-reversing data points constitutes a chunk of information in an individual's internal representation. Hence local curvature extrema may indicate boundaries in the perceptual grouping of connected lines thus creating numerous visual chunks for higher level cognitive processing. Interpreting long-term trends may therefore be difficult, because it requires integration of these visual chunks, which may require effortful cognitive processes such as spatial transformations.

If this is the case, language might be a useful tool to support spatial cognition. Evidence suggests that attending to spatial language when encoding visual scenes

can help construct representations that support spatial reasoning (Loewenstein & Gentner, 2005) and can influence memory of spatial scenes (Feist & Gentner, 2007). Furthermore, language can provide a user-goal during the study of a visual scene (i.e. a purpose for engaging with the scene), which may then activate relevant schema and guide visual-spatial attention (Brunyé & Taylor, 2009; Rothkopf, Ballard, & Hayhoe, 2007; Yarbus, 1967). Eye-tracking studies of relatively simple graphs indicate that visual attention appears to be driven by user-goals and graph knowledge (Carpenter & Shah, 1998; Peebles & Cheng, 2003) and hence using language to influence these top-down processes might help users to attend to and encode appropriate information in time-series line graphs.

The aim of Experiment 1 was to characterize difficulties, if any, in trend interpretation by asking participants to look at and then describe a real-world time-series graph that contained an underlying long-term trend as well as substantial short-term variability. Experiment 2 then asked whether a linguistic warning, providing an interpretation strategy, might improve encoding of long-term trends.

Experiment 1

To see if people correctly identify long-term trends from time-series graphs that also show significant short-term variability, verbal descriptions were collected from individuals exposed to a real-world graph showing such characteristics. The graph chosen (Figure 1) shows data for Northern Hemisphere spring snow cover extent between 1922-2012, published by the Intergovernmental Panel on Climate Change (IPCC, 2013). The IPCC is an international scientific body tasked with communicating policy-relevant scientific information to policy makers. The Figure therefore has societal relevance. Furthermore, the data indicate a significant downward trend over the whole time-period, together with substantial inter-annual variability. The authors indicate that snow cover extent has decreased since the mid-20th century (IPCC, 2013), suggesting that this is an important communication goal.

Method

Participants Twelve undergraduate students (10 female, two male) from the University of East Anglia took part in the study in return for course credit or a nominal payment. Their average age was 21 years (range 19–29 years). None of the participants were studying environmental sciences.

Apparatus and Materials The stimulus was presented on a TFT LCD monitor (51cm x 29cm), set to 1280 x 720 pixels. Eprime Version 2.0 (Psychology Software Tools Inc., Sharpsburg, USA) was used to control stimulus presentation and record data. Verbal responses were captured via a headset microphone. The stimulus consisted of Figure SPM.3a from the IPCC Summary for Policy Makers (IPCC, 2013) (Figure 1).

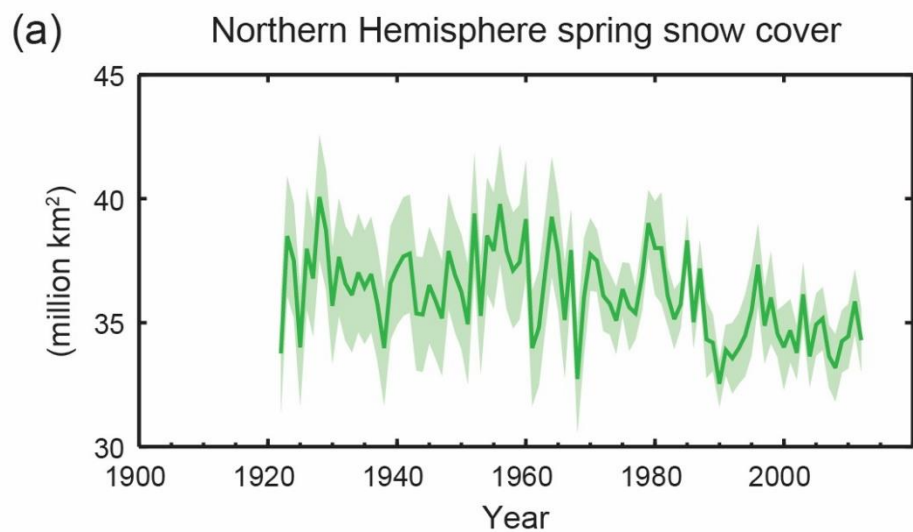


Figure 1: SPM.3a from Figure SPM.3: Multiple observed indicators of a changing global climate (IPCC, 2013).¹

Procedure The figure was presented for 15 seconds – during this time, participants were asked to simply look at the figure. They then saw a ‘Now describe’ prompt and the same figure re-appeared on the screen, at which point participants were asked to describe what they thought it was trying to show. The figure remained on screen until the participant completed their verbal response, up to a maximum time limit of 45 seconds.

Coding Verbal descriptions were coded to assess the presence (1) or absence (0) of the following aspects: (a) the data represent changes in snow cover over time; (b) a general downward trend; (c) a downward trend between ~1960 and ~2012; (d) short-term variability/fluctuation.²

Results and Discussion

All twelve participants correctly identified that the data represented changes in snow cover over time, but only five participants (42%) described a downward trend over the whole data. One of these participants also described a downward trend between ~1960 and ~2012. Of the five participants who described either type of downward trend, one also described the short-term variability (20%), but of the seven participants who did not describe either downward trend, five described the short-term variability (71%) ($p=.01$, Fisher’s Exact Test). These pilot data suggest that when presenting graphs that contain an underlying long-term trend and substantial short-term variability, spontaneous interpretation of the long-term trend may be far from guaranteed.

Experiment 2

The pilot data from Experiment 1 indicate that the long-term trend may not be readily interpreted in graphs that also show substantial short-term variability. The aim of Experiment 2 was therefore to test whether a linguistic warning that provides a strategy for interpreting long-term trends (by ignoring task-irrelevant

features) would improve encoding of the long-term trend; and if so, whether this is driven by changes in visual attention (measured using eye tracking). In addition, Experiment 2 investigated whether reducing, or removing intermediary x-axis tick marks and labels might have a beneficial effect on the encoding of long-term trends, as their presence might cue people to read-off data values or focus on short-term (inter-tick/-label) trends.

Method

Design To test spatial representations of the long-term trend (i.e. gradient) and short-term variability (i.e. amplitude), a forced choice task was employed in which participants were shown a graph to study and then asked to make a ‘same’ or ‘different’ judgment on a following test graph. The test graph was either identical to the study graph (same); had the same peaks and troughs as the study graph but with a different gradient (gradient different); had the same gradient as the study graph but with exaggerated peaks and troughs (amplitude different); or was completely different to the study graph (completely different). The number of x-axis ticks, either 2, 5 or 9, was varied across each type of test graph (see Figure 2 for examples).

To test the effect of a linguistic warning on cognition of the graph, participants were randomly allocated to either receive a warning asking them to ignore extreme values in order to consider the long-term trend (warning), or to receive no such warning (no warning). The experiment was therefore a 4 (trial type) x 3 (x-ticks) x 2 (warning) design, with trial type and x-ticks as within participant variables and warning as a between participant variable.

Participants Forty undergraduate students (29 female, 11 male) from the University of East Anglia took part in the study in return for course credit or a nominal payment. Their average age was 21 years (range 18-30 years).

Apparatus A Tobii TX300 Eye Tracker (Tobii Technology AB, Danderyd, Sweden) with integrated TFT LCD monitor (51cm x 29cm) set to 1280 x 720 pixels was used for stimulus presentation and collection of eye gaze data at 300Hz. Eprime Version 2.0 (Psychology Software Tools Inc., Sharpsburg, USA) was used to control stimulus presentation and record data. Responses for same-different trials were given using the ‘Z’ and ‘M’ keyboard keys. Response key mappings were reversed and counterbalanced between warning conditions. Verbal responses were recorded via a headset microphone. Eye gaze data were analyzed using OGAMA Version 4.5 (A. Voßkühler, Freie Universität Berlin, Germany), using default parameters for fixation detection.

Linguistic Warning The linguistic warning was displayed in 28pt Calibri and read: “WARNING When looking at graphs, people are often misled by extreme data points – short-term fluctuations in the data can obscure the long-term trend. To avoid errors, it is useful to ignore extreme data points to correctly identify the long-term trend.”

Graph Stimuli Twenty-four study time-series graphs were created (1126 x 510 pixels), each plotting 17 data points. Graphs showed an underlying positive, negative or flat long-term trend. Data points for each graph were created by sampling residuals at random from a normal distribution, which were then applied to a baseline positive, negative or flat linear trend graph. The x-axis was labelled ‘Years’ and the y-axis was labelled either as ‘Medication use (doses)’, ‘Infections (patients)’, ‘Temperature (°C)’, ‘Rainfall (mm)’, ‘Income (GBP £)’, or ‘Expenditure (USD \$)’. The x-axis covered a range of 16 years, with the starting year always between 1900 and 1994. A caption was created for each graph that simply read ‘[variable] over time.’

A positive, negative and flat trend study graph was allocated to each trial type. A test graph was then created for each study graph. Test graphs for the same condition were identical to their corresponding study graph. Test graphs for the gradient different condition were created by a transformation of the study graph

that resulted in a visual rotation of the graph line by ± 2 degrees. Test graphs for the amplitude different condition were created by multiplying the residuals of the study graph by a factor of 1.4. Three new graphs were created to serve as test graph pairings for the completely different trials. For each study and test graph pairing, three variants were created, each showing 2, 5 and 9 x-ticks (Figure 2). The remaining study graphs were allocated to true-false and describe filler trials, which also included variations for each level of x-ticks.

Areas of Interest (AOI) AOIs were defined for each study graph by first determining a circle around each data point with a maximum diameter that would avoid overlapping adjacent data points (58 pixels). A parallelogram with height 58 pixels, width 1002 pixels (2.0 x 34.5 degrees of visual angle), was then fitted over the line of best fit of the graph data, determined by linear least squares regression. This formed the line of best fit AOI (6.3% of screen area). A convex hull was then determined around the outer edges of these shapes, which formed the whole data AOI (mean 22.1% of screen area). An extreme data AOI was defined as the area of the whole data AOI that sat outside of the line of best fit AOI (mean 15.8% of screen area) (Figure 3).

Procedure Participants were informed that the study was investigating how people understand line graphs and they then received instructions on screen before a practice block of trials. The eye tracker was calibrated and then participants in the warning condition received the warning on screen and were instructed to read it before starting the first of three blocks of trials. Participants in the no warning condition simply started the first block of trials after eye tracker calibration. Each trial consisted of a study phase (Figure 4) during which participants were asked to look at and study the caption and the graph. The caption was presented prior to the graph to help control time spent reading the caption. The study phase was followed by one of three task cues (Figure 4). For same-different trials, participants had to make a same-different judgment about a test caption and then about a test graph in comparison to the

study caption and study graph. Participants were instructed to give a response as quickly as possible when the caption/graph appeared.

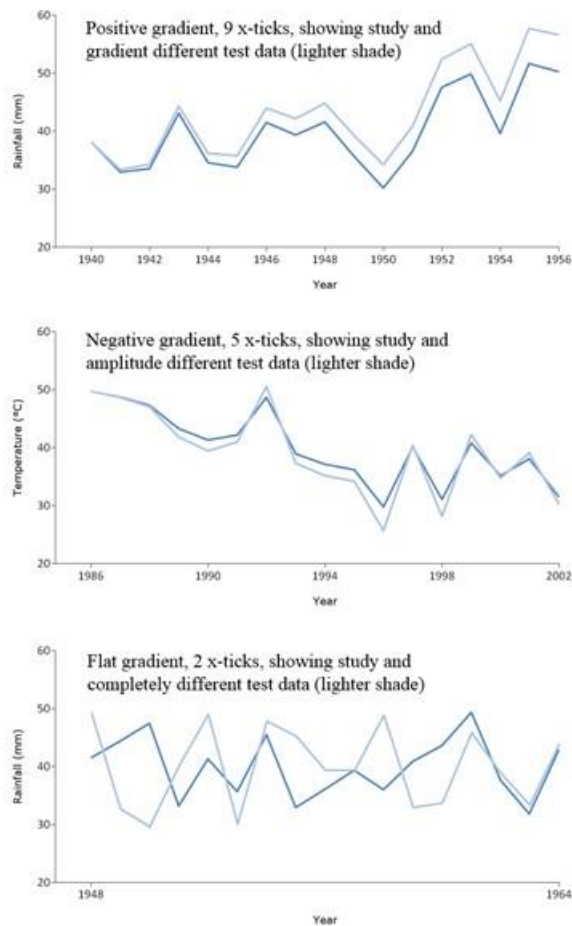


Figure 2: Three examples of the study and test graphs in Experiment 2.

Each block consisted of 12 same-different trials (three of each of the different trial types), presented in random order. Three true-false trials and three describe trials were included in each block to encourage participants to study the graphs in a naturalistic way and to ensure depth of encoding. Each x-tick variation of a given graph was presented in a different block. Blocks of trials were counterbalanced across participants and the eye tracker was re-calibrated at the start of each block. At the end of the third block, participants in the warning condition were asked what they remembered about the warning. The study lasted approximately 1 hour.

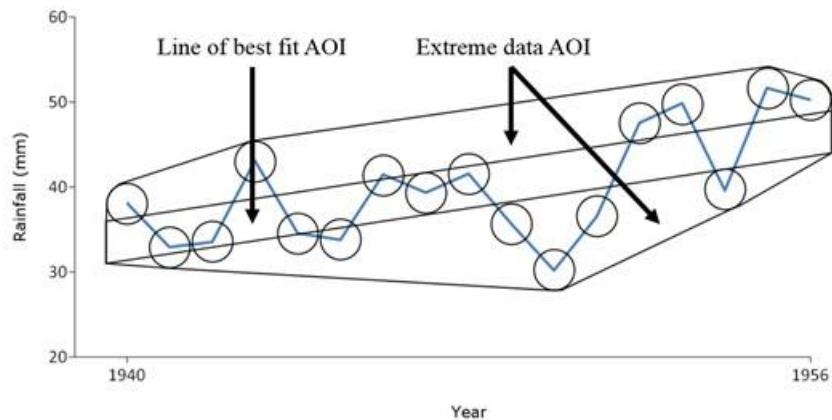


Figure 3: Line of best fit AOI and extreme data AOI for one of the 24 study graphs in Experiment 2.

Results and Discussion

Only same-different trials in which a correct response was given to the test caption and a response was given to the test graph were included in the analyses (i.e. trials in which participants correctly remembered the caption and then went on to make a judgement about the graph). Six participants were removed from further analyses: one participant who subsequently reported monocular vision impairment; one participant whose accuracy on completely different trials was 11% (lower than three *SD* from mean accuracy); and four participants in the warning condition who could not remember any detail about the warning when asked at the end of the study (and so may not have encoded it).

Task Performance Sensitivity to detect differences between the graphs of same-different trials was measured using d' in order to assess response accuracy with the effects of response bias removed. Participants' d' scores were analyzed with a 3 (trial type) x 3 (x-ticks) x 2 (warning) mixed ANOVA. There was a main effect of trial type, $F(2,64)=59.603$, $p<.001$, partial $\eta^2=.651$. Bonferroni post-hoc tests indicated a significant difference between amplitude different trials and

completely different trials ($p < .001$), and gradient different trials and completely different trials ($p < .001$), indicating that participants had a greater ability to detect differences between study and test graphs when the test graph was completely different, than when only the amplitude or gradient was different.

There was no main effect of x-ticks, $F(2,64)=0.504$, $p=.606$; and no main effect of warning, $F(1,32)<0.001$, $p=.994$. However there was a significant interaction between trial type and warning, $F(2,64)=3.459$, $p=.037$, partial $\eta^2=.098$ (Figure 5). Post-hoc examination indicated that participants in the no warning condition performed significantly worse on gradient different trials ($M = 0.251$, 95% CI ± 0.222) than amplitude different trials ($M = 0.667$, 95% CI ± 0.274) ($p=.008$), whereas those in the warning condition performed about equally on gradient different trials ($M=0.504$, 95% CI ± 0.293) and amplitude different trials ($M=0.479$, 95% CI ± 0.349). There was no significant x-ticks x warning interaction, $F(2,64)=3.041$, $p=.055$; and no three-way interaction, $F(4,128)=1.162$, $p=.331$, indicating that the number of intermediary x-ticks did not influence sensitivity to detect changes in the long-term trend.

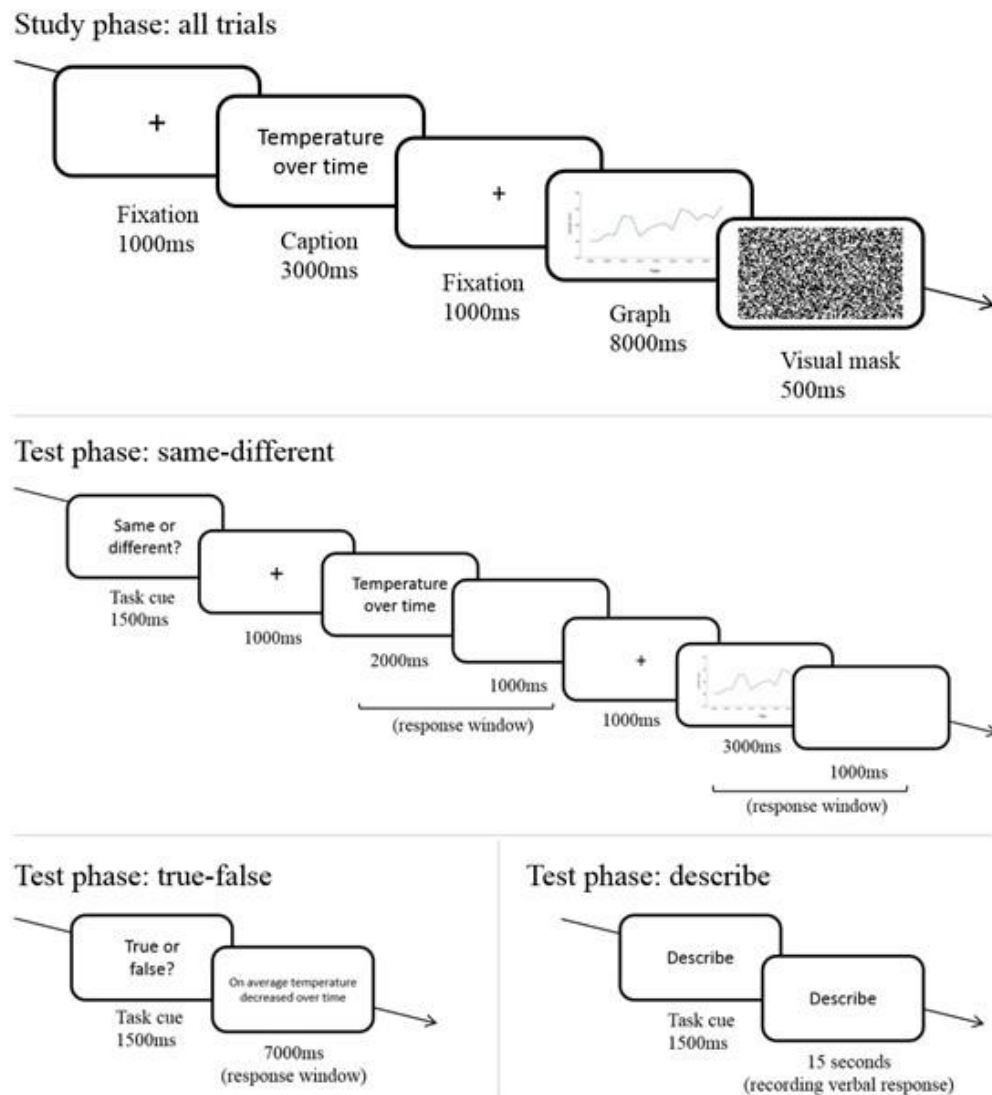


Figure 4: Presentation of same-different and filler trials.

Using language to provide task-relevant knowledge improved sensitivity to detect differences in task-relevant information (i.e. the long-term trend) relative to other information (i.e. amplitude). Furthermore, this did not appear to come at the expense of an impaired sensitivity to detect differences in the other information.

To investigate if the effect of the warning on gradient performance deteriorated over time, d' values were recalculated by collapsing data across x-ticks (as there was no significant x-ticks main effect or interaction), and then splitting out the data by block. A 2 (warning) x 3 (trial type) x 3 (block) mixed ANOVA was then performed. Results were consistent with the first mixed ANOVA, and there was

no three way interaction between trial type, warning and block, $F(2.903,92.895)=0.189, p=.898$ (with Greenhouse-Geisser correction), indicating that there was no evidence to suggest that the trial type x warning interaction was modulated by the duration between the warning and the block of trials. This suggests that the warning was encoded into long-term memory and applied throughout the study. These results indicate that the warning had a lasting effect on participants' judgements, suggesting that in the absence of explicit user-goals, using language to impart graph knowledge may direct subsequent interpretation of the data.

Visual Attention To investigate if the improved discriminability of the gradient found in the warning condition might be driven by differences in visual attention during encoding, fixation durations for the AOIs of the study graphs were calculated. Fixations were calculated for same-different trials in which a correct response was given to the caption and a response was given to the test graph, all true-false trials in which a response was given, and all verbal trials. Trials for four participants were excluded from further analysis as they had poor eye tracking calibrations. Individual trials were excluded if >15% of eye tracking samples were missing, or if there was a continuous period >700ms of data missing (10.7% of trials). As there was no main effect or interaction of x-ticks in the d' data, fixation data were collapsed across x-ticks.

At study, participants in the warning condition spent significantly longer fixating within the line of best fit area than participants who did not receive the warning, $t(19.802)=2.119, p=.024$ (one-tailed, equal variances not assumed) (Table 1). Conversely, there was no significant difference in total fixation duration of the extreme data area between the two groups, $t(25.137)=-0.352, p=.728$ (two-tailed, equal variances not assumed), nor a significant difference in total fixation duration in the whole data area, $t(28)=1.288, p=.208$ (two-tailed, equal variances assumed). Taken together, the task performance and visual attention results suggest that using language to provide graph knowledge can direct visual attention to task-relevant information during encoding, which then enables the

creation of a more detailed internal representation of the graphed data (rather than merely an alternative representation) and can influence subsequent interpretation.

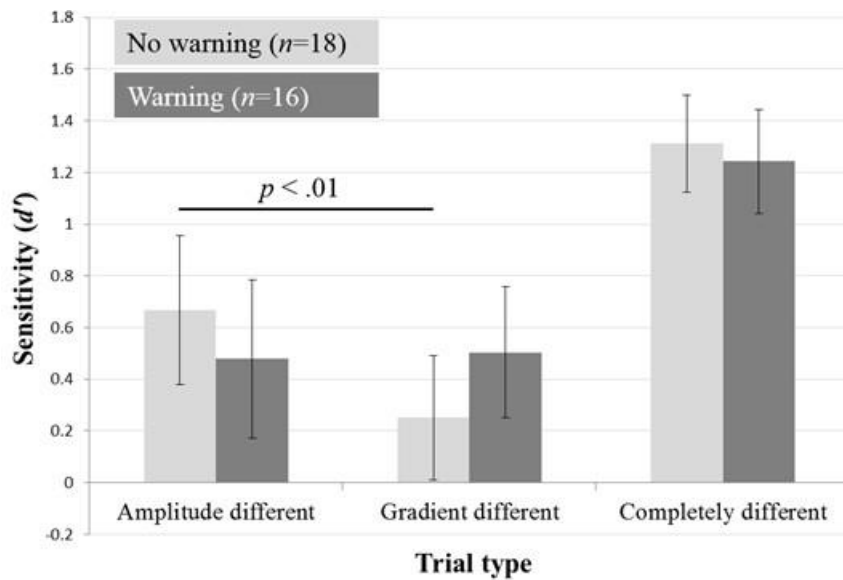


Figure 5: Average sensitivity (d') for each trial type and warning group, with 95% confidence intervals.

Table 1: Mean (M) and standard deviations (SD) of fixation duration in ms during study for each AOI.

Area of interest	No warning ($n=16$)		Warning ($n=14$)	
	M	SD	M	SD
Line of best fit	1426	(432)	1919	(772)
Extreme data	1587	(586)	1525	(356)
Whole data	3013	(884)	3444	(952)

General Discussion

The research presented here supports and builds on existing theoretical research on display comprehension and has important implications for communicators of time-series data. Pilot data from Experiment 1 found that interpretations of a real-world time-series line graph that contained a high degree of short-term variability (and therefore many trend reversals) did not elicit correct descriptions of the long-term trend in more than half of the participants. This is consistent with the hypothesis that trend reversals provide salient visual cues that break down connected lines into separate visual chunks, which may then be difficult to integrate into a representation of the long-term trend. Experiment 2 found that in the absence of an explicit user-goal or an interpretation strategy, users created better representations of the short-term variability than the long-term trend. However, when provided with an interpretation strategy via a linguistic warning, participants encoded both the long-term trend and short-term variability equally well.

In contrast to previous research investigating changes to the layout and format of a display in order to make task-relevant patterns explicitly represented (e.g. Shah, Mayer, & Hegarty, 1999), the research presented here highlights top-down cognitive processes on the identification and interpretation of data patterns. Language may be an effective way of providing graph knowledge, which can then be drawn on to direct visual attention to relevant visual features and support appropriate spatial inferences.

This may be especially pertinent when communicating complex data sets that contain several communication goals. For example, climate scientists may wish to communicate the long-term trends of indicators of a changing climate, as well as enabling individuals to understand that short-term variability in these indicators exists. Language may provide a useful tool to direct users to consider aspects that require complex inferential processes (such as the long-term trend) in addition to the salient patterns in the display. Given the need for individuals to interpret graphs to make informed decisions and play an active role in society, there is a

need to extend our theoretical understanding of display comprehension, and to apply and test out theoretical insights in real-world communication problems. The research presented here supports both of these aims.

Footnotes:

1 Multiple observed indicators of a changing global climate: (a) Extent of Northern Hemisphere March-April (spring) average snow cover. All time-series (coloured lines indicating different data sets) show annual values, and where assessed, uncertainties are indicated by coloured shading.

2 Inter-rater reliability across all aspects and all coding: $\kappa = 1.000$, $p < .001$.

Acknowledgments

JH was supported by a PhD studentship from the School of Psychology, University of East Anglia and a travel grant from the Spatial Intelligence & Learning Centre (SILC), Temple University. (SBE-1041707 from the National Science Foundation).

References

- Ali, N., & Peebles, D. (2013). The effect of gestalt laws of perceptual organization on the comprehension of three-variable bar and line graphs. *Human Factors*, 55(1), 183-203.
- Brunyé, T. T., & Taylor, H. A. (2009). When goals constrain: Eye movements and memory for goal-oriented map study. *Applied Cognitive Psychology*, 23(6), 772-787.
- Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied*, 4(2), 75-100.

- Carswell, C. M., Emery, C., & Lonon, A. M. (1993). Stimulus complexity and information integration in the spontaneous interpretations of line graphs. *Applied Cognitive Psychology*, 7(4), 341-357.
- Feist, M. I., & Gentner, D. (2007). Spatial language influences memory for spatial scenes. *Memory & Cognition*, 35(2), 283-296.
- Freedman, E. G., & Shah, P. (2002). Toward a model of knowledge-based graph comprehension. In M. Hegarty, B. Meyer, & N. H. Narayanan (Eds.), *LNAI 2317: Diagrammatic Representation and Inference*. Berlin, Germany: Springer.
- Glazer, N. (2011). Challenges with graph interpretation: A review of the literature. *Studies in Science Education*, 47(2), 183-210.
- Hegarty, M. (2011). The cognitive science of visual-spatial displays: Implications for design. *Topics in Cognitive Science*, 3(3), 446-474.
- Hoffman, D. D., & Richards, W. A. (1984). Parts of recognition. *Cognition*, 18(1), 65-96.
- IPCC (2013). Summary for Policymakers. In T.F. Stocker, D. Qin, G.-K. Plattner, et al. (Eds.), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK and New York, USA: Cambridge University Press.
- Kosslyn, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology*, 3(3), 185-226.
- Loewenstein, J., & Gentner, D. (2005). Relational language and the development of relational mapping. *Cognitive Psychology*, 50(4), 315-353.
- Peebles, D., & Cheng, P. (2003). Modeling the effect of task and graphical representation on response latency in a graph reading task. *Human Factors*, 45(1), 28-46.
- Rothkopf, C. A., Ballard, D. H., & Hayhoe M. M. (2007). Task and context determine where you look. *Journal of Vision*, 7(14), 16, 1-20.

- Schwert, G. W. (2011). Stock volatility during the recent financial crisis. *European Financial Management*, 17(5), 789-805.
- Shah, P., Mayer, R. E., & Hegarty, M. (1999). Graphs as aids to knowledge construction: Signaling techniques for guiding the process of graph comprehension. *Journal of Educational Psychology*, 91(4), 690-702.
- Subak, S. (2003). Effects of climate on variability in Lyme disease incidence in the northeastern United States. *American Journal of Epidemiology*, 157(6), 531-538.
- Yarbus, A. L. (1967). *Eye movements and vision*. New York, USA: Plenum Press.
- Zacks, J., & Tversky, B. (1999). Bars and lines: A study of graphic communication. *Memory & Cognition*, 27(6), 1073-1079.