Observed Strategies of Drawing Data Representations

Fiorenzo Colarusso[1[0000-0002-1953-5959]], Peter C-H. Cheng [1[0000-0002-0353-5955]], Grecia Garcia Garcia[1[0000-0002-7327-7225]], Daniel Raggi[2[0000-0002-9207-6621]] and Mateja Jamnik[2[0000-0003-2772-2532]]

[1]University of Sussex, UK
{f.colarusso, p.c.h.cheng, g-garcia-garcia}@sussex.ac.uk

[2] Univeristy of Camdridge, UK
{daniel.raggi, mateja.jamnik}@cl.cam.ac.uk

Abstract. This study considers the drawing strategies of four diverse participants as they copied a line-graph and a bar-chart. Video recordings of the transcriptions were analyzed stroke by stroke. Diverse global drawing strategies were used for the line graph whereas a similar approach was used by all on the bar-chart, but with local differences. The fluency of the participants' performance varied substantially, particular in viewing frequency of the stimuli. Differences in the strategies can be explained in terms of how they perceptually chunked the stimuli. Sample GOMS models were constructed in order to demonstrate verify that chunking explains the drawing strategies. The potential of using drawing transcription tasks to assess user's competence with graphs and charts is discussed.

Keywords: Chunking, competence measurement, diagrams, bar chart, line graph, graph comprehension, task analysis, GOMS, HCI, representations.

1 Introduction

How people draw diagrams has been rather neglected in the diagrams research literature, but it is worth studying for many reasons. It is an interesting complex cognitive phenomenon that is worth investigating its own right (von Sommers, 1984). A drawer's approach to drawing depends on their familiarity with the diagram and whether it is reproduced from long term memory, copied or traced (Obaildellah & Cheng, 2015). Different drawing strategies can reflect learners' understanding of technical topics (Roller & Cheng, 2014). Our particular reason for studying the nature of drawing is motivated by whether individual's diagram drawing behaviors reflect their familiarity with the diagrams being produced. Are there signals in drawing behaviour that can be used to assess an individual's competence with a particular class of diagrams? The focus here is on data charts and graphs.

In turn, our interest in assessing people's competence with particular diagrams, and representations more generally, is motivated by the *Rep2Rep* project that is attempting to build an automated system to selected appropriate representations for individual as the attempts to solve specific problems in some target domain (Jamnik & Cheng, 2021). Representation selection is essential because the choice of representations substantially determines the ease and success problem solving and learning. The Rep2Rep framework involves two aspects: (a) selecting representations that are formally adequate for the problem using a formal AI system and (b) picking representations that are cognitively suited to individuals. This paper focuses on the second. A key aspect of assessing the cognitive suitability of a representation for an individual is knowledge of their level of familiar with the given representation, that is how well the representation is understood. So, we need quick and reliable means of evaluate people's familiarity with different representations. We contend that when a individuals reproduce a particular representation, or copy a diagram, their behaviour will provide signals that reflect how familiar they are with the diagram from which measures of their competence can be derived. This paper makes a first step towards that aim by observing and modelling the variety of drawing strategies that people use to reproduce given some data diagrams. As background consider: (i) how the graphical structures of graphs and charts are critical to how they are understood in general; (ii) how the role of chunking in comprehension, and how the analysis of chunks provides means to assess competence; (iii) existing approaches to assess familiarity.

1.1 Interpreting and Comprehending Graphs

As our goal is to assess users' familiarity with data diagrams, we should consider what it means for someone to understand them. In general, line graphs are employed to depict x-y trends while bar charts support comparisons between bars that are closer together on the display. Wu et al. (2010) argue that the tendency to associate lines with trends is due to the cognitive naturalness and the ease of the perceptual process. Shah, Mayer & Hegarty (1999) found that format and scale influence the graph interpretation, thus bar charts should be used when two independent variables are equally important, while line graphs when a particular trend is more relevant. According to the model for graph interpretation (Shah & Carpenter 1995, Shah et al. 1999), three processes are particularly relevant: encoding, transition of visual features to conceptual relations, and referential processes. An accurate encoding of the major visual pattern in the graph, such as whether there is a straight or jagged line, is essential for the correct comprehension of the graph. The transition of visual features to conceptual relations requires the retrieval of quantitative knowledge associated with the visual pattern, such as the knowledge that a downwardly curved line represents a decreasing function. Therefore, when the visual pattern evokes familiar quantitative concepts the comprehension is effortless. Moreover, Peebles & Cheng (2003) found that minor changes in the graph design affect the user performance in graph reading task in terms of the visual pattern to find the required information. Thus, it is likely that competence in graph comprehension is closely linked to users' familiarity with the perceptual patterns or visual features of graphs and charts.

Process models of graph comprehension also imply the importance of grasping the organization and processing structure of graphs and charts. The Construction-

Integration (CI) model (Kintsch, 1988) constitutes an effective approach for graph comprehension. CI model states that comprehension can be subdivided into two sub-phases: construction phase and comprehension phase. Moreover, three pools of units are included in the model: visual features, domain knowledge, and interpretation propositions (Freedman et al. 2002). During the construction phase, textual information, prior knowledge, and goals interact to form a coherent representation of the available information. During the comprehension phase when the information is depicted in the graph by visual features and can be linked with prior knowledge without making inferences, the comprehension is effortless while, if inferences are needed the process becomes effortful. Similarly, Hegarty (2005) proposed a Model of display comprehension to explain how people construct a mental model starting from the display visualization. The model claims that bottom-up information (design features) interact with top-down processes (prior knowledge). Thus, familiarity and background knowledge may influence the manner in which attention is directed to the external display and how information is perceived, interpreted, and modeled internally (Kriz & Hegarty, 2007). As explained by Freedman and colleagues (2002), taking as an example a line graph, an expert can integrate into a coherent mental representation the visual features and the interpretation of data while a novice, lacking the prior knowledge of the graph, can't explicit represent information, thus inferences are necessary and the comprehension becomes effortful. Thus, an expert automatically forms a link between the visual features (the shape of the line) and the theoretical interpretation of the data. When a perceiver lacks the relevant prior knowledge, or the display does not explicitly represent information that must then be inferred, comprehension is effortful. If diagrams do not contain all the information that a user needs to use, thus familiar users, due to their prior knowledge, can interact better than novices with specific diagrammatic representation, as their background knowledge guides them in information processing and inference process, compared to generating, by the graph, interaction new knowledge and awareness (Cheng el al., 2001).

1.2 Chunking in Competence Measurement

As graphical features and structure underpin comprehension theories from psychology and cognitive science, this can be recruited for the explanation of graph competence in a manner that enable methods for its assessment to be developed. Of particular relevance are schema (Bartlett, 1932) and chunking (Miller, 1956) theories. Chunking can be defined as a process by which new chunks are made to improve the learnability of the information, so the greater someone's familiarity the richer the content of their chunks in memory. According to Gobet et al. (2001), state the chunking process has a dual nature based on two opposite assumptions: the first which defines chunking as deliberate and conscious (goal-oriented chunking) while the second as an automatic and continuous process that occurs during the perception (perceptual- chunking). Thus, tasks involving processes that require perceptual processing and deliberate processing of chunks will be determined by the individual's personal organization of chunks in their memory, so there will be rich signals in the behaviors that reflect the structures of those chunks (Gobet, 2005; Gobet & Simon, 1996; Gobet, at al., 2001; Holding, 1985). For instance, the time between successive task activities, pauses, varies depending on where in the hierarchy of chunks processing is occurring. Pauses between the production of intra-chunk elements (within a chunk) will be short, whereas the pauses between actions spanning inter-chunk boundaries (between chunks) will be relatively long. Further, the higher in the hierarchy of chunks an inter-chunk transition occurs, the greater will be the pause.

Based on these ideas about pause analysis of chunk structures, Cheng and colleagues (2014; Cheng & Rojas-Anaya, 2007; Albehaijan and Cheng, 2019) developed an approach to assess competences using transcription tasks, in which stimuli, such as mathematical formulas or program code, are copied. The measure of competence in those tasks exploit a strong and robust temporal signal that reflects the structure of chunks in an individual's memory. In particular, pauses between successive written characters are sensitive to chunk structure, so the shape of the distribution of pauses varies with the competence of the transcriber. Demonstrations of the potential of assessing competence using temporal chunk signals in transcription tasks have established various measures that are well correlated with independent measures of competence.

The issue is now whether this approach can be used to assess individuals' competence with diagrammatic representations. The previous work has mainly focused on linear symbolic notations and natural language, the key issue is whether the technique and measures are applicable to diagrams? Diagrams are 2D, do not have an obvious linear format to follow during transcription, but some previous works on chunking in diagrams drawing suggest that there is some potential (Cheng et al. 2001; Obaidellah & Cheng 2015; Roller & Cheng 2014). The diagrams targeted in that works did not include data graphs and charts. So, the question for this paper is whether clear signs of chunking are manifested in the transcription of charts and graphs. Specifically: (a) Will the transcription of these representations show temporal signals, patterns of pauses between drawing actions, that reflect the structure of chunks? (b) Will those signals varying between individuals in ways that suggest they possess different chunk structures?

These questions will be addressed empirically and theoretically. In the next main section, a small-scale study of four participants transcribing diagrams is presented, which answers the questions affirmatively. In the third main section, the task analysis – using GOMS – is used to model the differences in the observed behavioural strategies in order to show that the distribution of pauses can be attributed to the possession by the participants of different hierarchical chunk structures.

1.3 Existing Methods for the Assessment of Graph Familiarity

To end the introduction, two previous approaches to the assessment of familiarity with graphs should be acknowledged. Xi (2005) assessed graph competence for line graphs and bar charts using a Graph Familiarity questionnaire. This questionnaire has verbal statements, which are judged on a 6-point scale, in groups concerning: participants' prior experience using graphs, their ability to read graphs, and their typical reactions to graphs. Moreover, both in bar charts and line graph, Xi found that planning time affects the accuracy of the verbal graph description as participants captured the major points of the graph and described more elements.

The other approach by Cox & Grawemeyer (2003) assesses how people organize their knowledge of external representations (ERs) through a card-sorting task. They found that expert ability using ERs in reasoning and problem solving was associated with high performance in semantic distinction and accurate naming of ERs, thus high competence participants produced few categories in the ER card-sorting task as they had better mental representations of ER knowledge, and perceived the semantic commonality between visually different ERs.

2 Observing Strategies of Drawing Graphs and Charts

It is imaginable that participants might differ little in how they transcribe graphs and charts, because such diagrams have been designed to make particular visual features and structures particularly salient. This might mask any effects of familiarity with these representations and hence behavioural signals due to chunking. This would contrasts with linear sentential notation in which the structure of expressions depend heavily on the content of the expressions. Thus, it is essential to show that the transcription behaviors of these 2D representations do reveal signs of chunking. To this end, a small-scale observation experiment was conducted.

2.1 Experiment

Participants. Four right-handed participants with Master's in different subjects were recruited. All completed Xi's (2005) graph familiarity questionnaire (on a scale of 1 to 6, where 6 is high familiarity). Their scores (and subjects) are: P1=4.7 (Finance); P2=4.3 (Engineering); P3=2.9 (Literature); P4=2.3 (Law). The scores are clearly dichotomous and consistent with the participants' educational speciality.

Materials. Fig. 1 shows the two stimuli used. In order to improve the task difficulty, we used a grouped bar graph (Fig. 1b) from the Wall Street Journal, "Auto Industry, at a Crossroads, Finds Itself Stalled by History", January 2, 2006. We designed the line-graph (Fig. 1a) specially so that it had two sets of points that might be perceived as corners of two hexagons, as potential distractor to the three data lines. Each was accompanied by a summary of their general meaning. To show the stimuli, we used a laptop computer running a logging program specially written in our lab. We recorded the participants' drawing from above with a video-camera. All drawing actions, pen strokes, were coded using the ELAN video analysis software (Sloetjes & Wittenburg, 2008) and the duration of pauses between strokes computed with milliseconds (ms) accuracy.

Procedure. The experiment consisted of two trials where participants copied the stimulus on a blank sheet of paper using the participant-driven "hide-show" interaction method (Albehaijan & Cheng 2019), in which stimulus only appears on the computer screen when the participant holds down a special key. To write on the sheet participants must release the key and the stimulus is hidden. This method allowed us to record: (a) view-numbers – the total number of views of the stimulus in a trial; (b) view-times – the duration of each look at the stimulus; (c) writing-times – the time spent writing between two successive views.

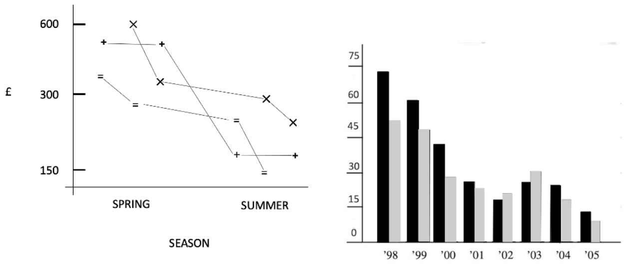

Fig. 1. (a) Line graph and (b) bar chart employed as stimuli for the study

The results for each are presented and discussed separately and then followed by some general discussion.

2.2 Line Graph: Results

Observing the drawing strategies exhibited by the participant during the task, we found various approaches across the participants, including: P2 & P3 – reproduce each set of data, switching continuously between dots and lines; P4 – reproduce each set of data in turn, with a tendency to do all the data points first followed by the connecting lines; P1 - drawing all data points first, for the two hexagons, then fill in the connecting lines. Consistent with previous studies, all participants had distributions of pauses (times between strokes) that appear to reflect hierarchal organization of chunks memory (Cheng & Rojas-Anaya, 2005; Roller & Cheng, 2014; Thompson et al. 2017). Specifically, the pauses be for the first stoke of meaningful groups of elements is longer than the pauses within those groups: longer pauses seem to reveal with the

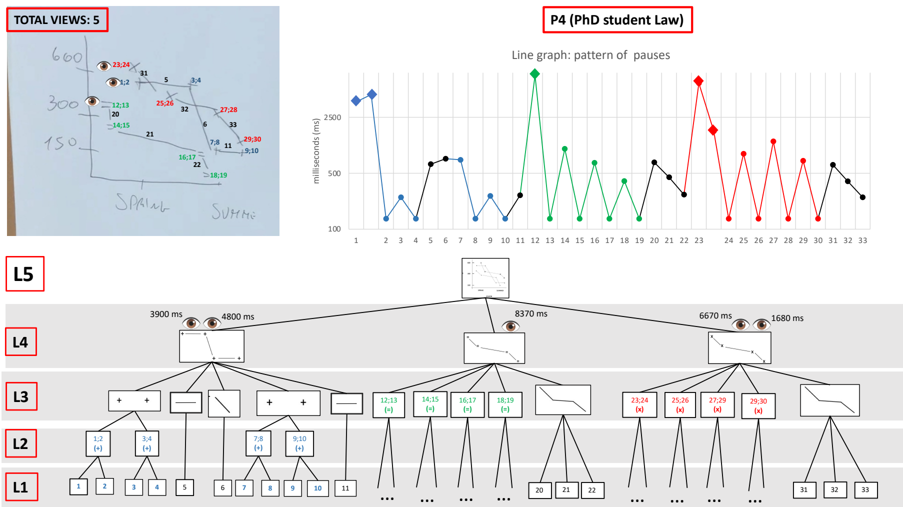

Fig. 2. P4's (a) drawing, (b) pause profile, and (c) chunk process hierarchy.

transitions between chunk or sub-chunk boundaries and may reflect the total amount of cognition required prior to each pen-stroke.

Take, for example, the strategies employed by P4, Fig.2, and P1, Fig.3, as representative of those with high (P4) and low (P1) familiarity with graphs and charts. Each figure is organized in three sections: (a) the drawing produced by the participant with stokes numbered; (b) a pause profile line graph for pauses (log scale) for each consecutive stroke, with diamond dots which correspond to the participant's views of the stimuli; (c) a chunk process hierarchy tree diagram derived from the pause profile. Level L5 is the whole drawing. L4 is for chunk(s) acquired by a view(s) of the stimulus. L3 is a sub-chunk level, where new sub-chunks are defined by a pause threshold of 500 ms (Obaidellah & Cheng, 2015; Roller & Cheng, 2014). The overall structure of the trees changes little with reasonable variations of the threshold. L2 and L1 are levels for symbols and strokes.

P4's graphical production, Fig. 2, appears to be organized by a graphic schema that separates datapoints and connecting lines, particularly in the second and third chunks, where all the datapoints are produced before the line connecting them is completed. Each schema is acquired in one view (or two consecutive views) and the connecting lines appear to be treated as sub-chunks. In contrast, P1's production, Fig. 3, has a different strategy, that starts with datapoints at the extremes of the plot, then completion of points within each hexagon, and finally the three sets of connecting lines. The profile of pauses shows less evidence of large chunks, but still includes signs of chunks. Consequently, the process hierarchy is shallower as the sub-chunk level is absent (L3). P1 has a higher number of views than P4, and initially appears to be treating the three sets of data points as a single field.

P2 and P3 present a similar overall approach focusing on each data set in turn, like P1, but they took approximately twice the number of views (P2=8; P3=9). At a lower level, however, they broke down each dataset into groups of a few points and lines, each associated with a view. Thus, it appears they did not use a high-level schema for each dataset, but were nevertheless chunking.

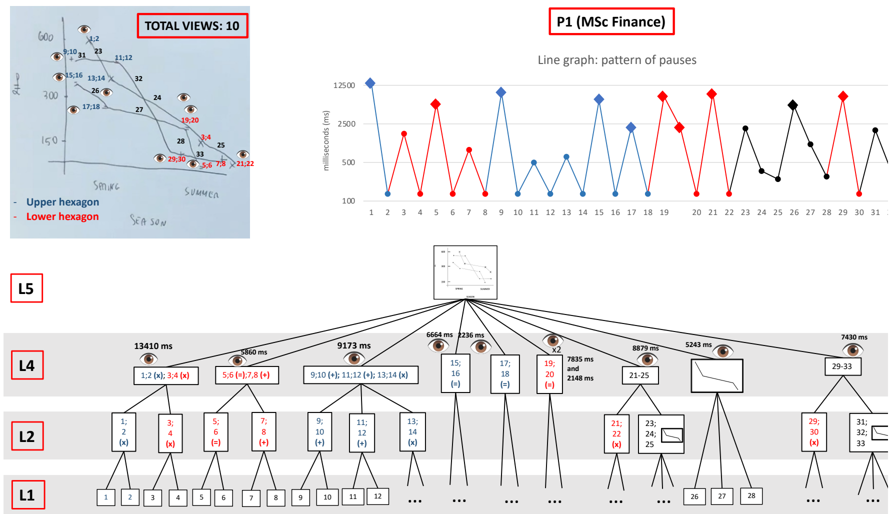

Fig. 3. P1's (a) drawing, (b) pause profile, and (c) chunk process hierarchy

2.3 Line Graph: Discussion

Diverse stategies were employed by the participants during the task as expected for such a heaviliy perceptually oriented task (van Sommers, 1984). The strategies vary overall in relation to the datasets and also locally in relation to subgroup of elements within a data set. Despite these differences, it is clear from the pause profiles and structure of the derived processing hierarchies – which incorporates information about the occurrence of views beyond the pause threshold – that chunking is being used in the transcription of the line graph.

Contrary to expectations, the two participants with greatest familiarity with graphs and charts, were not exploiting chunking processes the most in their drawings. Only P4, the lowest scorer on the questionnaire, can be characterized as showing processes with a clear pattern of chunking and sub-chunking. However, P1, who scored highest on the questionnaire, adopted an approach that might be considered as an attempted to use an overly sophisticated strategy, because he ignored the three distinct groups of data, for some reason; faced with a large field of points, he appears to have tried to demark their overall shape and then fill in the individual points. This may be linked to Roller & Cheng's (2014) and Obaidellah & Cheng's (2015) observations that the drawing of complex diagrams may follow a decomposition strategy in which an overall frame is first produced and followed by details within. Despite this strategy, there is evidence of chunking, albeit not consistently associated with the sets of data and hence the intended meaning of the line graph.

2.4 Bar chart: Results & Discussion

Unlike the line graph, where evident differences occur in the drawing strategies, for the bar chart all the participants shared similarities in terms of their drawing sequences. The overall strategy adopted by all was to reproduce the bars from left to right. Also, the black bar was always drawn first in each pair of bars.

At a lower level, evidence of chunking is apparent in the number of views required by each participant and their pause profiles. The number of views varied markedly between participants: P1=10, P2=16, P3=6, P4=8 views. As there are 8 pairs of bars (Fig. 1b), P1, P3 and P4 require approximately one view per couple of bars whereas P2 dealt with one bar at a time. Given the similarity of the overall strategy, we were able to derive a general pause profile per each participant, shown in Fig .4 (numbers 1 to 5 indicate the pause sequence). All the values are means, except the 1^{st} of each participant and the and the 4^{th} of P2 for which the first quartile (Q1) of the view time was used to predict the pause duration and note the log scale. The pause at the start of each bar is longer than that for strokes within a bar and the pause for the start of the second bar is shorter than first, but longer than the pauses within the first bar. Thus, all participants are shown signs of treating each pair of bars as a chunk and each bar within as a sub-chunk.

It is odd that P2, who had a high score familiarity questionnaire, was not grouping two bars into one chunk. After the trial P2 explained that his overall goal was to accurately represent the values of the bars, which accounts for his one bar at a time approach. As P2 was attending to extra information, it is possible that this may have sufficiently loaded his working memory that no capacity was spare for the second bar. The approach is reflected in his generally longer durations of pauses for the first line of each bar compared to the other participants. Despite this difference in strategy, it is noteworthy that his pause profile has the same overall shape as the others.

The pause profiles are independent of the precise order of the drawing of lines

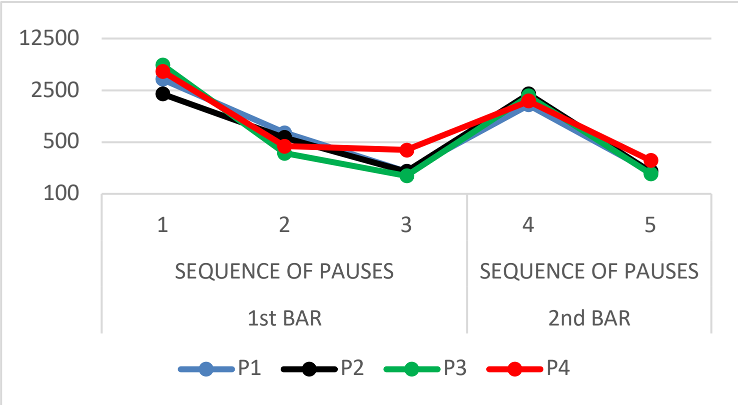

Fig. 4. Participants' pause profiles for one pair of bars in the bar chart.

within the bars. P1, P2 and P3 started the black bar drawing: (i) left line, (ii) top line and (iii) right line. However, they differ for the direction from which they start the left line: P2 drew the line from the top to the x-axis whereas P3-P4 did contrariwise. P4 drew bars differently producing in sequence: (i) top line, (ii) left line and (iii) right line. The production of the second bar was consistent for all participants. Despite the marked difference in the specific strategy of line production, it is clear that this is secondary to the role of chunking in their performance as shown by the profile of pauses in Fig. 4.

2.5 Overall Discussion of the Experiment

The purpose of the experiment was to investigate the possibility that the task of transcribing graphs and charts could be used as a basis of a method to assess users' familiarity or competence in particular classes of representation. In particular, it is essential to show that chunking has a major role in the production process, so that measures of chunking can be engaged, such as the number of views and distributions of pauses (Albehaijan & Cheng, 2019; Cheng, 2014). Overall, there is clear evidence of chunking in participants' transcriptions of both the line graph and bar chart stimuli. This is demonstrated by the pause profiles and also the coherence of derived chunk process hierarchies, that is the putative chunks correspond to meaningful groups of elements (Figs. 2, 3, 4). Chunking provides good explanations of the participants' performance despite the wide variety of strategies they used, at global and local levels. Further, the size of derived chunks (2-4 sub-chunks) is in line with chunking theory for complex tasks. In this respect the transcription of graphs and charts may have potential as method for competence measurement.

However, the diverse drawing strategies are problematic as they may not be associated with chunks that are related to the meaning of the target representation but encode superficial visual features. As noted by Kriz and Hegarty (2007), the interaction between prior knowledge and the bottom-up features presented by a stimulus influences the perceptual processing and therefore the way in which chunks are made and drawn. The sequence of production may be affected by a wide range of perceptual factors, intrinsic the design of our stimulus, inviting the subjects to adopt a specific order of production (Van Sommers, 1984). Strong visual patterns, such as those highlighted by gestalt principles of visual perception, may determine a drawing strategy. This is a particular concern, when such patterns coincide with meaningful features of the data being displayed, such as smooth trends in the data. For instance, are P2 and P3 drawings of each trend of line (where they switch continuously between dots and lines) uninfluenced by the "Law of Continuity", where the line is perceived as continuous movement in order to minimize abrupt changes? Alternatively, is P4's graphic schema of grouping visually similar units together influenced by the principle of "Grouping by Similarity"?

An implication of this is that methods must be developed so that the strategies adopted during transcription are closely tied to whatever meaningful chunk the participants have of the target stimuli; that is, chunks must reflect the way in which information is encoded when transcribing a representation rather than accidental perceptually salient patterns. At minimum, participants must be instructed that precise values of data points are not of concern, in order to prevent behaviors such as P2's narrow precision goal on the bar chart. We might, for instance, instruct participants to focus on the meaning and communicative intent embodied by the representations.

Although only four participants contributed transcriptions for each representation, it is noteworthy that despite the clear difference in the familiarity of the two pairs of participants, there was no indication of difference in competence. One explanation of why is that any effect of familiarity may have been masked by the issue of diverse drawing strategies, some of which might have been perceptually driven. Another explanation is that the selected stimuli (Fig. 1) are too simple for the selected participants. In future work we will test more complex line graphs and bar charts.

3 CPM-GOMS Verification of the Line Graph Chunking

The aim of this second part of the study is to obtain converging evidence that the behaviors in the transcription of the graph and chart were largely determined by chunking processes. We will use cognitive modelling, specifically task analysis, for this and focus specifically on the line graph. The idea is to generalize the strategies used by participants on the line graph to produce an ideal chunk hierarchy that shares the common features of the individual approaches (Fig. 6, top). This ideal chunk hierarchy is then adopted as the basis for building a task analytic model composed of a typical sequence of cognitive processes for similar tasks and standard values of timings for basic cognitive operations. The chunk hierarchy determines sequencing of operators in the model. From the model we derive pauses preceding each stroke in the simulated drawing of the ideal chunk hierarchy. If the profile of pauses of the model matches the typical distribution of the participants' pauses, this implies that chunking processes are also responsible for the participants' pauses.

3.1 CPM-GOMS Modelling

We adopt the GOMS approach to task analysis. John & Kieras (1996) GOMS is a family of modelling techniques that analyses the user complexity of interactive systems. Each type of GOMS task analysis consists of a hierarchical task decomposition based on the Goal, Operators, Methods, and Selection rules (Card et al. 1983). The Goal is what the user is trying to accomplish on several levels of abstraction. Operators are atomic elements that generally hold a fixed execution time. Methods consist of set of operators commonly applied to achieve a goal. Selection rules choose between methods. The Cognitive, Perceptual, Motor GOMS (CPM-GOMS) technique, using the Model Human Processor (MHP) as a framework (Card et al., 1983), is the most suitable for drawing transcription tasks because it can deal with parallel execution of visual perception, cognitive and motor operations. In CPM-GOMS, the perceptual processor is responsible for transforming external information into a form that the cognitive system can process; the cognitive processor uses contents of WM and LTM to make decisions and schedule actions with the motor system; and, the motor processor is responsible to translate thoughts into actions. The CPM-GOMS architecture employs the PERT/Gantt-like charts to represent the relations between the operators and the critical path that derives from it allows to estimate the total time required for the task execution. We used the software Cogulator (Estes, 2016) to model the drawing task.

Fig. 5 shows a section of the Cogulator model for the ideal chunk hierarchy.

In this section we will explain how the model was built and we will show a short template used for CPM-GOMS model (Fig. 5). The model has three principal types of statements: GOAL, .Goal and .Also. The GOAL statements represent the main goals required to perform the transcription task, specifically a perceptual goal and a drawing goal. The .Goal statements are sub-goals included within the main goal and deal with the different items that must be drawn. Also is used to represent the parallel processes that occur during the pauses when the pen is moving or hovering between the inscriptions. Each operation has a separate line in the code. The number of full stops before a code word indicates the nesting level of the operator.

Most of the values employed for the operators are provided by the literature (John & Newell, 1989; Gray & Boehm-Davis, 2000). At the beginning of each chunk within the drawing GOAL, we assumed a recall operator of 1200 ms (John et al. 1989; Lee, 1995) to retrieve information from WM. For the .jump and .pen_stroke motor operations, which execute pen moves between strokes and the strokes themselves, we picked values based on the average times of the participants for .pen_stroke operators while we predicted the jump durations summing the values for all cognitive operators that we assumed occur in parallel during the pause. Moreover, as pauses between symbols' inscriptions are assumed to be automatic processes, are not included in CPM-GOMS analysis.

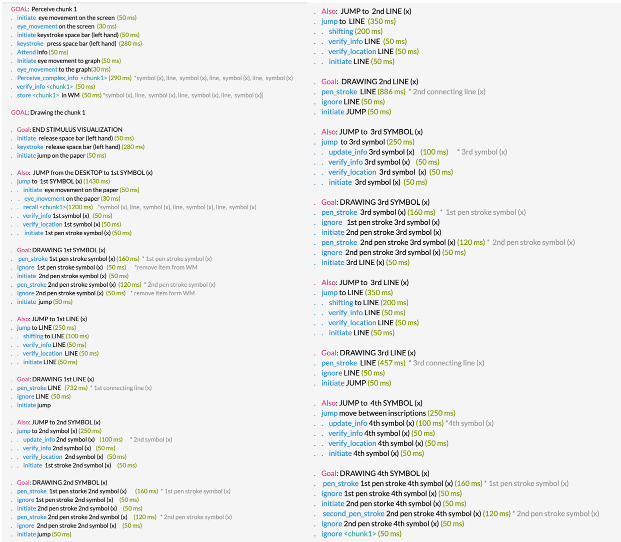

Fig. 5. Example of Cogulator code for a CPM-GOMS for the drawing of the first chunk for the idealized model (fig.6). Two main GOAL are required, one for perceiving the stimulus and the latter for drawing the chunks. The second is always broken down into several sub-goals (i.e. .Goal statements) necessary for drawing the elements within the chunk (e.g. symbols, connecting lines). All the pauses between the inscriptions represented by the .jump motor operators, were obtained summing the cognitive operators values which occur in parallel within .Also statements.

As diagram drawing is not typically modelled by GOMS, we decided to deal with the spatial information separately as the spatial coordinates values are fundamental both in drawing and graph comprehension. So, some non-standard operators are defined: verify_location, shifting and updating. We decided to assign 50 ms to verify_location, to match that of the standard GOMS verify_information. Furthermore, as the task is complex, it likely involves executive functions (EF) (Miyake et al. 2000, 2012; Morra & Panesi, 2016). EF operators comprise those mental capacities necessary for formulating goals, planning how to achieve them, and carrying out the plans effectively (Lezak, 1982), they differ from the cognitive functions (CF) as they explain how and whether a person goes about doing something, rather than what and how much. Thus we also define EF cognitive operators: ignore, shifting, and updating. The shifting operator is a main component of the cognitive flexibility. It is an ability used by people to represent their knowledge about a task and the possible strategies in which to engage (Cañas et al. 2006). A shifting operator of 100 ms occurs when, during drawing, a participant switches between necessary sub-tasks to accomplish a sub-chunk (e.g., before drawing each connecting line following a symbol) while its value increase to 200 ms during the switching between sub-chunks as an upper item in the hierarchy need to be picked (e.g. the transition from pen stroke 5 to 6 in fig.6). An .updating operator of 100 ms is required before drawing each point or line which represent the data points respectively in the line graph and bar charts while an .ignore operator occurs whenever a pen stroke is made before starting the subsequent cognitive operation.

3.2 Modelling Results

The idealized chunk hierarchy has three chunks with two or three sub-chunks, Fig. 6. Applying standard sequences of CMP-GOMS operators to this hierarchy, with the values given above, the full series of operations needed for the task and their timings were assembled. Pauses between the end of each stroke and beginning of each stroke were computed and the pause profile graph plotted, Fig. 6 bottom.

The overall shape of the profile resembles the profiles for the participants (e.g., Figs. 2 & 3). There are long pauses for the views, very short pauses for strokes withing symbols, but critically medium and short pauses for sub-chunk, which are comparable to the participants durations of pauses. When the idealized chunk structure is modified, for example to more closely reflect a specific participant, the precise pattern of the pause profile also changes, but the overall nature of the distribution remains the same. Thus, the match between the CPM-GOMS models and that of the participants, suggests that chunking is primarily responsible for the shape of the profiles, and hence chunking is critical in these drawing transcription tasks.

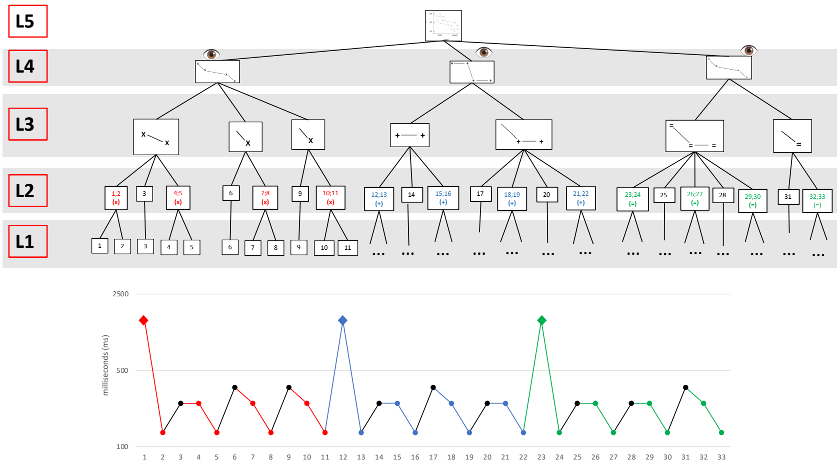

Fig. 6. Idealized chunk hierarchy (top) and predicted pause profile (bottom).

4 Discussion

Chunking is one of the most studied phenomena in cognitive science due to its ubiquity in learning and information processing. We aim to produce a method to assess competence in representations that goes beyond current tests that only indirectly assess familiarity using questionnaire, verbal descriptions or a simple task (Cox & Grawemeyer, 2003; Xi 2005). Critical for our approach is to show that chunking occurs in transcriptions involving the drawing of diagrams, so the previously established measures of chunk-based assessment can be adopted (Cheng et al. 2014; Cheng & Rojas-Anaya, 2007; Albehaijan & Cheng, 2019).

In the observations of drawings of the four participants on the line graph and bar chart, evidence was found of chunking. Pauses between strokes had distributions typical of tasks involving chunking, and values typical of chunking. Longer pauses for inter-chunk transition at higher level chunks and shorter pauses for intra-chunk transitions at lower levels. From the pause profile putative chunk hierarchies were systematically derived, and they exhibited structure typical of chunking processes.

Generalizing over the predicted chunk hierarchies an idealized chunk hierarchy was constructed and used as the foundation of a CMP-GOMS task analytic model. The good correspondence between the model and participants pause profiles, particular in the levels and magnitudes of pauses, adds weight to the claim the chunking was central in the transcription processes. This suggests, at least in principle, that such drawing transcription task may have potential as measures of competence.

However, the diversity of strategies and the actual patterns of drawn elements suggest that the chunks may often reflect obvious perceptual patterns and conventions rather than chunks and schemas that underpin the participants underling knowledge of the two representations. Thus, recommendations for refinements to the method have been suggested (in Section 2) to ensure that meaningful chunks are most likely to be probed.

Acknowledgements

We thank Gem Stapleton, from Cambridge University, for her comments and suggestions for this paper. This work was supported by the EPSRC grants EP/R030650/1, EP/T019603/1, EP/R030642/1, and EP/T019034/1.

References

1. Albehaijan, N., & Cheng, P. C.-H. (2019). Measuring programming competence by assessing chunk structures in a code transcription task. In A. Goel, C. Seifert, & C. Freksa (Eds.), Proceedings of the 41st Annual Conference of the Cognitive Science Society (pp. 76-82). Austin, TX: Cognitive Science Society
2. Bartlett, F. C. (1932) *Remembering: a study in experimental and social psychology*. Cambridge, UK: Cambridge Univer. Press. P. 329.
3. Bengio, .Y. Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. IEEE transactions on pattern analysis and machine intelligence, 35(8):1798–1828, 2013.
4. Bower, G. H. (1970) Organizational factors in memory. Cognitive Psychology, 1(1), 18-46. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/0010028570900034.
5. Brewer, W. F. (1999) Schemata. In R. A. Wilson & F. C. Keil (Eds.), MIT encyclopedia of the cognitive sciences. Cambridge, MA: MIT Press. Pp. 729-730.
6. Canas, J. J., Fajardo, I., & Salmeron, L. (2006). Cognitive flexibility. In W. Karwowski (Ed.), International encyclopedia of ergonomics and human factors (2nd ed., pp. 297–301). Boca Raton, FL: CRC Press.
7. Card, Stuart, Moran, Thomas P., and Newell, Allen, The Psychology of Human-Computer Interaction, Lawrence Erlbaum Associates, Hillsdale, NJ (1983)
8. Chase, W. G., & Ericsson, K. A. (1982). Skill and working memory. The Psychology of Learning and Motivation, 16, 1–58.

9. Chase, W. G., & Simon, H. A. (1973). Perception in chess. Cognitive Psychology, 4, 55–81.
10. Chase, W. G., & Simon, H. A. (1973). Perception in chess. Cognitive Psychology, 4(1), 55–81.
11. Cheng, P. (2020). A Sketch of a Theory and Modelling Notation for Elucidating the Structure of Representations. Diagrams (2020)
12. Cheng, P. C. H., Lowe, R. K., & Scaife, M. (2001). Cognitive science approaches to understanding diagrammatic representations. Artificial Intelligence Review, 15(1–2), 79–94.
13. Cheng, P. C.-H. (2014). Copying equations to assess mathematical competence: An evaluation of pause measures using graphical protocol analysis. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (pp. 319-324). Austin, TX: Cognitive Science Society.
14. Cheng, P., & Rojas-Anaya, H. (2007). Measuring mathematic formula writing competence: An application of graphical protocol analysis. Proc. of the Thirtieth Annual Conference of the Cognitive science Society.
15. Cheng, P., McFadzean, J., & Copeland, L. (2001). Drawing out the temporal signature of induced perceptual chunks. Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society, 200–205.
16. Cheng, Peter C-H (2016). What constitutes an effective representation? In: Jamik, M., Uesaka, Y. and Schwartz, S. (eds.) Diagrammatic Representation and Inference: 9th International Conference, Diagrams 2016. LNAI (9781). Springer, Heidelberg, Germany, pp. 17-31.
17. Cox, R., & Grawemeyer, B. (2003). The Mental Organisation of External Representations. Proceedings of EuroCogSci 03, 91–96.
18. Egan, D. E., & Schwartz, B. J. (1979). Chunking in recall of symbolic drawings. Memory & Cognition, 7(2), 149–158.
19. ELAN (Version 6.0) [Computer software]. (2020). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan
20. Estes, Steven. (2016). Introduction to Simple Workload Models Using Cogulator.
21. Freedman, E. G., & Shah, P. (2002). Toward a model of knowledge-based graph comprehension. LNCS, 2317, 18–30.
22. Freedman, E. G., Shah, P. S. (November 2001). Individual differences in domain knowledge, graph reading skills, and explanatory skills during graph comprehension.
23. Gobet, F. (2005), Chunking models of expertise: implications for education. Appl. Cognit. Psychol., 19: 183-204.
24. Gobet, F., & Simon, H. A. (1996). Templates in chess memory: A mechanism for recalling several boards. Cognitive Psychology, 31, 1–40.
25. Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C. H., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. Trends in Cognitive Sciences. https://doi.org/10.1016/S1364-6613(00)01662-4
26. Gray, W. D., & Boehm-Davis, D. A. (2000). Milliseconds matter: an introduction to microstrategies and to their use in describing and predicting interactive behavior. Journal of Experimental Psychology: Applied, 6(4), 322-335.
27. Gray, W. D., & Fu, W. T. (2004). Soft constraints in interactive behavior: The case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. Cognitive Science, 28(3), 359–382.
28. Harnishfeger, K. K., & Pope, R. S. (1996). Intending to forget: The development of cognitive inhibition in directed for- getting. Journal of Experimental Child Psychology, 62(2), 292–315.

29. Holding, D. H. (1985). The psychology of chess skill. Hillsdale, NJ: Erlbaum.
30. Jamnik, M., & Cheng, P. C.-H. (2021). Endowing machines with the expert human ability to select representations: why and how. In S. Muggleton & N. Chater (Eds.), *Human-Like Machine Intelligence*. Oxford: Oxford University Press.
31. John, B. E., & Kieras, D. E. (1996). The GOMS family of user interface analysis techniques. ACM Transactions on Computer-Human Interaction, 3(4), 320–351.
32. John, B. E., & Newell, A. (1989). Cumulating the science of HCI: From S-R compatibility to transcription typing. Conference on Human Factors in Computing Systems - Proceedings, May, 109–114. https://doi.org/10.1145/67449.67472
33. Kintsch, W. (1988). The role of knowledge in discourse comprehension. A construction-integration model. Psychological Review, 95, 163-182.
34. Kriz, S., & Hegarty, M. (2007). Top-down and bottom-up influences on learning from animations. International Journal of Human Computer Studies, 65, 911–930
35. Lee, A. (1995). Exploring User Effort Involved in Using History Tools Through MHP/GOMS: Results and Experiences. 109–114. https://doi.org/10.1007/978-1-5041-2896-4_18
36. Lezak, M.D. (1982). The problem of assessing executive functions. International Journal of Psychology, 17, 281-297.
37. Lohse, J. (1991). A cognitive model for the perception and understanding of graphs. Conference on Human Factors in Computing Systems - Proceedings, 137–144.
38. Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 63(2), 81–97.
39. Obaidellah, U. H., & Cheng, P. C. H. (2015). The role of chunking in drawing Rey complex Figure. Perceptual and Motor Skills, 120(2), 535 555.
40. Palmer, S. (1977) Hierarchical structure in perceptual representation. Cognitive Psychology, 9(4),
41. 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.
42. Raggi D., Stapleton G., Stockdill A., Jamnik M., Garcia G. G. and Cheng P. C.-H., "How to (Re)represent it?," 2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI), Baltimore, MD, USA, 2020, pp. 1224-1232.
43. Roller, R., & Cheng, P. C.-H. (2014). Observed strategies in the freehand drawing of complex hierarchical diagrams. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (pp. 2020-2025). Austin, TX: Cognitive Science Society.
44. Shah, P., & Carpenter, P. A. (1995). Conceptual limitations in comprehending line graphs. Journal of Experimental Psychology: General, 124(1), 43–61.
45. 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.
46. Sloetjes, H., & Wittenburg, P. (2008). Annotation by category - ELAN and ISO DCR. In: Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)
47. van Sommers, P. 1984. Drawing and Cognition: Descriptive and Experimental Studies of Graphic Production Processes. Cambridge, UK: Cambridge University Press
48. Wu, P., Carberry, S., Elzer, S., & Chester, D. (2010). Recognizing the intended message of line graphs. LNCS, 6170 LNAI, 220–234.
49. Xi, X. (2005). Do visual chunks and planning impact performance on the graph description task in the SPEAK exam? In Language Testing (Vol. 22, Issue 4).