Well-matchedness in Euler and Linear Diagrams

Gem Stapleton[1,2][0000−0002−6567−6752], Peter Rodgers[2][0000−0002−4100−3596], Anestis Touloumis[3][0000−0002−5965−1639], and Andrew Blake[3][0000−0001−5856−4544]

[1] University of Cambridge, Cambridge, UK, ges55@cam.ac.uk
[2] University of Kent, Canterbury, UK, {g.stapleton,p.j.rodgers}@kent.ac.uk
[3] Centre for Secure, Intelligent and Usable Systems, University of Brighton, Brighton, UK
{a.touloumis,a.l.blake}@brighton.ac.uk

Abstract. A key feature of diagrams is well-matchedness, referred to as iconicity in philosophy. A well-matched diagram has a structural resemblance to its semantics and is believed to be an effective representation. In this paper, we view well-matchedness as a feature of diagrams' meaning carriers – syntactic relationships that convey meaning. Each meaning carrier may or may not structurally resemble, i.e. be well-matched to, its semantics. This paper provides the first empirical study that evaluates the impact of well-matched meaning carriers on effectiveness in Euler diagrams and linear diagrams. There are two key take-away messages: using only well-matched meaning carriers led to the best task performance and using both well-matched and non-well-matched meaning carriers in a single diagram should be approached with caution.

Keywords: Well-matched · Iconicity · Diagrams · Visualization · Sets.

1 Introduction

The notion of well-matchedness encapsulates the property of a diagram's syntactic relations corresponding, structurally, to its semantics [10, 4]. A highly related notion is the concept of iconicity: Peirce took iconicity to embody the structural resemblance of a syntactic entity (a sign) to its semantics (object) [13]. In this paper, we demonstrate that well-matchedness is a property of meaning carriers [15]. This paper presents empirical studies that evaluate the impact of well-matched meaning carriers on effectiveness in Euler diagrams and linear diagrams, with the study designs embodying this fine-grained view of well-matchedness. The studies use stimuli that vary the use of meaning carriers to convey information within diagrams to test the impact of well-matchedness on task performance.

Section 2 illustrates the key ideas of meaning carriers and well-matchedness and makes the first contribution of this paper: identifying well-matchedness as a property of meaning carriers, not global properties of diagrams. Section 3 evaluates well-matchedness in Euler diagrams, presenting the hypotheses to be tested, the design of the empirical study, the methods used to analyse the collected data and the results. Section 4 covers linear diagrams, proceeding in the same manner as section 3. We conclude in section 5. Supporting material can be found at www.eulerdiagrams.com/wellmatched.

2 Well-matchedness and Meaning Carriers

A fundamental aspect of any notation, diagrammatic or otherwise, is how it combines basic syntactic elements to form meaningful expressions. A meaning carrier is a relationship between syntactic elements that conveys either true or false information [15]. A key goal is to provide general theories about the relative cognitive effectiveness of competing diagram choices through understanding meaning carriers and their role in cognition. Meaning carriers allow us to identify information that is explicitly conveyed by a representation of information: this explicit information is defined to be observable from the representation [15]. It is also vital to understand meaning carriers when exploring well-matchedness: if a meaning carrier resembles the semantics it conveys then it is considered to be well-matched. Well-matchedness is a property of some meaning carriers but not others, distinguishing it from the notion of observability, and is hypothesised to help explain the relative cognitive benefits of competing diagram choices.

2.1 Meaning Carriers

Euler and linear diagrams both exploit spatial relationships between curves and, respectively, lines to convey information about sets. Each of these notations can be augmented with shading [7, 6] to convey information in a syntactically different way. We are only focusing on subset or disjointness relationships between pairs of sets and, therefore, are only concerned with the meaning carriers identified below. In general, other meaning carriers arise.

Meaning Carriers in Euler Diagrams Fig. 1 contains four diagrams that show information about the countries visited by people. We focus on part of the information conveyed: *everyone who visited Ukraine also visited Romania.* This is subset information: the set of people who visited Ukraine is a subset of those who visited Romania. In the leftmost diagram, only spatial relationships between circles convey information: the inclusion of one circle inside another is a meaning carrier, since it conveys information about the relationship between the corresponding sets. The first and second diagram in Fig. 1 both place Ukraine inside Romania, expressing the subset-style statement spatially using circles.

The third and fourth diagrams do not exploit an equivalent spatial relationship: Ukraine is not inside Romania. To convey the subset information, the region inside the former but outside the latter is shaded (shading identifies set emptiness). Shading can be viewed as an annotation that the corresponding set is empty: it is fundamentally different to spatial relations between circles when used to convey information. In the third diagram, therefore, the placement of shading inside parts of the Ukraine circle ensures that the diagram expresses

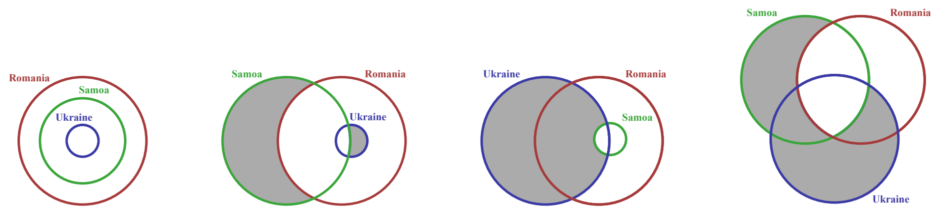

Fig. 1. Euler diagrams with varying meaning carriers: subset information.

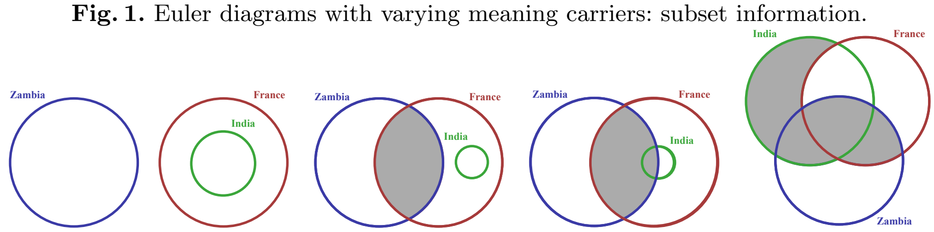

Fig. 2. Euler diagrams with varying meaning carriers: disjointness information.

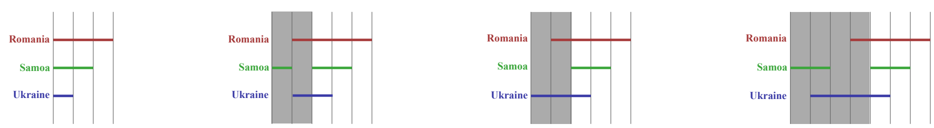

Fig. 3. Linear diagrams with varying meaning carriers: subset information.

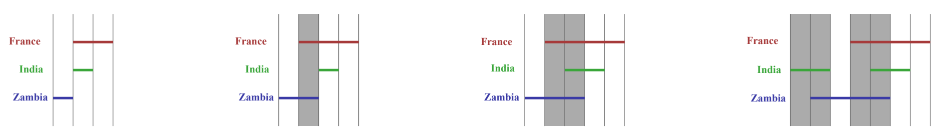

Fig. 4. Linear diagrams with varying meaning carriers: disjointness information.

that everyone who visited Ukraine also visited Romania. Similar reasoning can be applied to the fourth diagram.

Regarding disjointness relations, using Fig. 2 we focus on the statement *no one visited both India and Zambia*. In the first and second diagram, the non-overlapping nature of the India and Zambia circles is a meaning carrier expressing this information. By contrast, the third and fourth Euler diagram express the disjointness information by shading the region inside both India and Zambia.

Meaning Carriers in Linear Diagrams In Fig. 3, the linear diagrams convey the same information as the Euler diagrams in Fig. 1. In the first linear diagram, only spatial relations between lines convey information: if the x-coordinates of one line are entirely subsumed by those of another line then the set represented by the former is a subset of the latter. So, one line being completely overlapped by another is a meaning carrier. Hence, since (the line for) Ukraine is completely overlapped by Romania, the leftmost linear diagram expresses that everyone who visited Ukraine also visited Romania. The second diagram in Fig. 3 also ensures that the line for Ukraine is completely overlapped by the Romania line.

In the third and fourth diagrams, Ukraine is not completely overlapped by Romania. Here, meaning is conveyed using shading: in a shaded overlap[4], the represented set is empty. Thus, to express one set is a subset of another, we can shade the overlaps that include a line for the former but not the latter. In the third diagram, the overlap that includes the Ukraine line but not the Romania line is shaded. In the fourth linear diagram, the two overlaps that include Ukraine but not Romania are shaded. So both these diagrams express that everyone who visited Ukraine also visited Romania. As with the Euler diagram case, to extract the information that everyone who visited Ukraine also visited Romania relies on the presence of shading, not simply the spatial relationships between lines.

Regarding the expression of disjointness relations between sets, linear diagrams exploit either spatial relations between lines or shading. Using Fig. 4, we again consider the statement *nobody visited both India and Zambia*. In the first two diagrams of Fig. 4, the non-overlapping nature of the India and Zambia lines is a meaning carrier expressing this information. By contrast, the third and fourth linear diagram express the disjointness information by shading the overlaps that contain lines for both India and Zambia.

2.2 Well-Matchedness of Meaning Carriers

Recall that a meaning carrier is well-matched to its semantics if there is a structural resemblance between the way in which the meaning carrier expresses information and the information being expressed. Well-matchedness is a property of meaning carriers, not a global property of diagrams. To study well-matcheness in general, not just for Euler and linear diagrams, we need to identify the meaning carriers that are present in diagrams and whether they are well-matched. This fine-grained view of well-matchedness is potentially important for our continued study of the efficacy of diagrams.

Well-matchedness in Euler Diagrams Spatial meaning carriers arising from circles are well-matched. In the subset case, the inclusion of circle A inside B matches the meaning that all of set A is included in set B. Likewise, the disjoint interiors of two non-overlapping circles, C and D, matches the meaning that the two represented sets are disjoint. That is, in Euler diagrams, meaning carriers arising from circles are well-matched to their semantics. By contrast, there is no structural resemblance of shading to its meaning: the presence of a syntactic device – shading – is being used to express the absence of elements. Thus, meaning carriers arising from the use of shading are not well-matched. Returning to Figs 1 and 2, we can see that in both cases the leftmost diagrams only exploit well-matched meaning carriers, the middle two diagrams blend well-matched and non-well-matched meaning carriers and the rightmost diagram only uses non-well-matched meaning carriers.

[4] Overlapping lines represent set intersections with distinct overlaps are separated by vertical grid lines. The first diagram in Fig. 3 has three overlaps, with the first one representing the intersection of the three sets since all three lines appear.

Well-matchedness in Linear Diagrams Spatial meaning carriers arising from lines are also well-matched. In the subset case, line A being completely overlapped by line B matches the meaning that all of set A is included in set B. This is because the semantics are derived from the x-coordinates occupied by the line A forming a subset of those for the line B. So, the subset of x-coordinates at the syntactic level matches the subset of elements at the semantic level. Likewise, the non-overlapping nature of two lines, C and D, matches the meaning that the two represented sets are disjoint. That is, in linear diagrams, spatial meaning carriers arising from lines are well-matched to their semantics, just as for Euler diagrams. Again, there is no structural resemblance of shading to its meaning: shading is not well-matched. Returning to Figs 3 and 4, we can see that in both cases the leftmost diagrams only exploit well-matched meaning carriers, the middle two diagrams blend well-matched and non-well-matched meaning carriers and the rightmost diagram only uses non-well-matched meaning carriers.

2.3 Research Questions

The specific research questions addressed for these two notations are:

(RQ1) Do diagrams with only well-matched meaning carriers significantly improve performance over diagrams with some non-well-matched meaning carriers?
(RQ2) Do diagrams whose meaning carriers are well-matched to *the information to be extracted in a given task* significantly improve performance over diagrams that are *not* well-matched to *the information to be extracted*?

Answers will shed new light on the role of the well-matchedness of meaning carriers in Euler and linear diagrams and will inform the design of visual modes of communication: if well-matched meaning carriers yield demonstrable performance benefits then we should favour visualization methods that exhibit them. Moreover, if non-well-matched meaning carriers negatively impact performance then they should be avoided.

3 Evaluating Well-Matchedness in Euler Diagrams

To begin our study of well-matchedness, we will derive hypotheses concerning its role in Euler diagrams and its potential impact on cognition, measured via task performance. For our purposes, a representation is judged to support more effective information extraction than another if there is a significant accuracy or speed benefit. The evaluation of Euler diagrams was run in alongside the study on linear diagrams, presented in section 4: data were collected concurrently.

3.1 Hypotheses

The above discourse on the role of meaning carriers in conveying information and their potential to be well-matched leads to our first hypothesis:

[H1] to identify a piece of information from a diagram that is conveyed *using a well-matched meaning carrier* is *significantly easier* than identifying it *using a non-well-matched meaning carrier.*

This suggests that, in each of Figs 1 and 2, the two diagrams on the left will support significantly more accurate or, else, significantly faster time performance than the two diagrams on the right. What other differences between the diagrams in these figures might we expect to establish, empirically, if the well-matchedness of meaning carriers is a fundamental property that impacts task performance? To get a more precise handle on this we appeal to the theory of boundary segregation [5], which states that colour hue is favoured over shape when segregating boundaries, and the Gestalt Principles of Perceptual Organisation, in particular the principle of good continuation [17].

Suppose we wish to extract the information that *everyone who visited Ukraine also visited Romania* and that *no one visited both India and Zambia* from the diagrams in Figs 1 and 2 respectively. In each of the diagrams, their circles can be visually segregated from each other, primarily because of their distinguishing hues as colour[5] is more salient than form [5]. In addition, in each of the leftmost diagrams, the visual salience of the circles is further promoted: since no pair of circles have intersections between their boundaries, each circle exhibits the principle of good continuation. In the remaining diagrams, visual segregation is impaired because at least one pair of circles exhibit changes in good continuation at the points where circles intersect. Indeed, these changes in good continuation arise precisely because a non-well-matched meaning carrier is used and they promote the visual saliency of the intersection. These insights support [H1] and suggest that the leftmost diagram is more effective than the second diagram in each figure, leading to another hypothesis:

[H2] to identify a piece of information from a diagram that *only has well-matched meaning carriers* is *significantly easier* than identifying it from a diagram that *blends well-matched and non-well-matched meaning carriers, and expresses the desired information in a well-matched way.*

We suggest as the number of changes in good continuation increases (due to the use of non-well-matched meaning carriers), the less salient the circles become and the more difficult the task may get. However, we must also consider the crucial role of shading. That is, to understand the role of meaning carriers in information extraction, we need to understand the relative salience of circles and shaded regions. Since the same colour hue is used throughout for the shading, we posit that no one shaded region is more prominent than another, but the circles are more readily distinguishable due to their varying hues. Hence, the most salient information in an Euler diagram may arise from its well-matched meaning carriers. This further supports [H1] since it distinguishes the second and third diagrams in each of our figures: the second diagram uses a well-matched

[5] The discussion in this section is assuming the viewer of the representations is not impeded by colourblindness.

meaning carrier to convey the required information whereas the third diagram does not. Moreover, the third diagram uses a well-matched meaning carrier to express different information, which we speculate will act as a distraction from the task of identifying that *everyone who visited Ukraine also visited Romania* in Fig. 1 and that *no one visited both India and Zambia* in Fig. 2. It is known that, in general, syntax which causes a distraction from the target syntax required for the task can lead to reduced performance [9]. Applying this to Euler diagrams, the saliency of the well-matched meaning carrier in the third diagram of each figure inhibits the identification of the target, non-well-matched, syntax that must be interpreted to extract the aforementioned statements. By contrast, in the fourth diagram, there are no (salient) well-matched meaning carriers to distract form the task of identifying the required information. We obtain a third hypothesis:

> [H3] to identify a piece of information from a diagram that *uses only non-well-matched meaning carriers* is *significantly easier* than identifying it from a diagram that *uses both well-matched and non-well-matched meaning carriers and expresses the desired information using a non-well-matched meaning carrier*.

To summarise, combining [H1], [H2], and [H3], we expect the diagrams to be ranked, in terms of their ability to support the extraction of the stated information, as: the leftmost diagram is most effective ([H1] and [H2]), followed by the second diagram ([H2]), then the fourth ([H1] and [H3]) and, finally, the third diagram ([H1] and [H3]).

3.2 Methods

We recruited participants using the Prolific Academic crowdsourcing platform. Participants were asked to perform 8 tasks, presented in the *performance phase* of the study which was preceded by a *training phase*. Each task was a multiple choice question with five options, exactly one of which was the correct answer. There were two *preference phase* questions, one for subset-style statements and one for disjointness-style statements. The performance phase and preference phase each included additional questions to establish whether participants were paying attention. This is standard technique when crowdsourcing [8]. Data from inattentive participants – i.e. those who fail to answer at least one attention checking question correctly – are not included in any statistical analysis.

The study adopted a within group design. The participants would be asked, in the performance phase, a multiple choice question and were required to identify which of five options was correct. Two options were subset-style statements, two were disjointness-style statements and the fifth option was 'none of the above'. There were four tasks for subset-style statements and four for disjointness-style statements. All diagrams included in the paper to illustrate the study design are scaled for space reasons (typically to 30%). For the study materials, see www.eulerdiagrams.com/wellmatched.

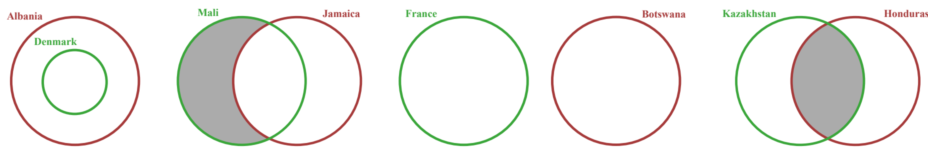

Fig. 5. The first four training Euler diagrams.

Training Phase Participants were shown diagrams similar to, but distinct from, those used in the other two phases. The first four training diagrams each displayed two sets. Of these, the first two conveyed subset information and the second two conveyed disjointness information. The fourth and fifth diagrams each used three sets, the first focusing on subset training and the second on disjointness. Fig. 5 shows the first four training diagrams, covering the use of spatial relations between circles and shaded regions as meaning carriers for subset and disjointness information. The training diagrams were presented in a fixed order.
Performance Phase This phase included four subset-style tasks and four disjointness-style tasks alongside one question to check for attentiveness. Fig. 6 shows a subset task where the answer is well-matched (correct answer: option 4) alongside the performance-phase attention checker. The options for the attention checker indicated which option to pick and, for the remaining options, used country names that did not appear in the diagram. The four tasks associated with each task type covered the following treatments:

– Well-Matched (WM): the diagram only exploits well-matched meaning carriers (spatial relationships between circles).
– Well-Matched to the Answer (WMA): the diagram exploits a well-matched meaning carrier to convey the correct answer, but also uses a non-well-matched meaning carrier (shading) to convey other information.
– Not Well-Matched to the Answer (NWMA): the diagram exploits a non-well-matched meaning carrier to convey the correct answer, but also uses a well-matched meaning carrier to convey other information.
– Not Well-Matched (NWM): the diagram only exploits non-well-matched meaning carriers.

Based on our hypotheses, we would expect our treatments to be ranked as WM $>$ WMA $>$ NWM $>$ NWMA, where $>$ means more accurate or faster.

The four Euler diagrams for the subset-style task conveyed the same information, up to label swapping, and varied only by their use of spatial relationships between circles and shading. Similarly, this was the only variation in the diagrams used for the disjointness tasks. Figs 7 and 8 show all of the diagrams used in the study. In each case, the diagrams are ordered (from left to right): WM, WMA, NWMA, NWM. No pair of diagrams shared a country name and, within each diagram, each country name started with a different first letter to reduce the potential for misreading. The colours assigned to the circles were derived from ColorBrewer to ensure they were perceptually distinct and suitable for categorised data [11].

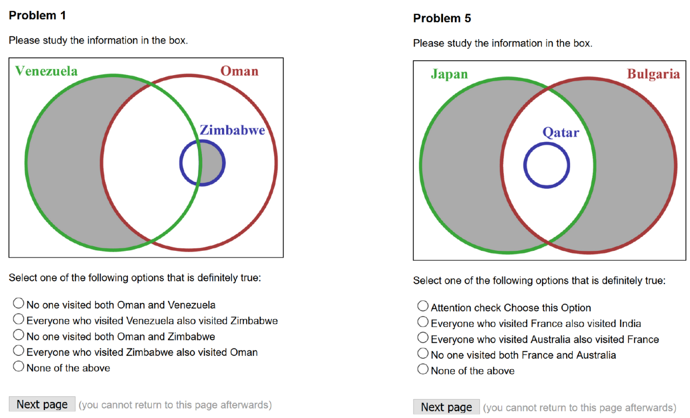

Fig. 6. Task presentation (left) and an attention checker (right).

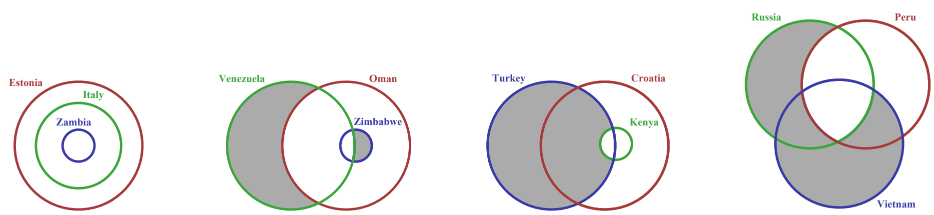

Fig. 7. The four Euler diagrams used for subset tasks.

Regarding the five options, the first four included two subset-style statements and two disjointness-style statements. The three incorrect options, excluding 'none of the above', had the sets involved randomly selected whilst ensuring that the options were not true. Regarding the correct answer, it would not be sensible to always place it in the same position (eliminating answer position as a variance across treatments): it would be easy to spot that the correct answer was always in, say, position 2. Table 1 indicates the positions of the correct answers for each statement style and treatment. In addition, we indicate the sets involved in the correct answer, abbreviating their names to first letters only and expressing the statement in mathematical notation (note the answers were always written as 'Everyone ... ' and 'No one ...' statements). For each participant, the order of the tasks was randomly generated, except that the attention checker always appeared after the fourth performance-phase task.

Preference Phase Participants were presented with two preference questions, asking them to rank four diagrams according to which most effectively conveyed a specified subset and, respectively, a disjointness statement. For the subset-style statement, the diagrams in Fig. 1 were used and the statement was *Everyone who visited Ukraine also visited Romania*. For the disjointness-style statements, the diagrams in Fig. 2 were used and the statement was *No one visited both India and Zambia*. The diagrams used in the preference phase were identical to the eight

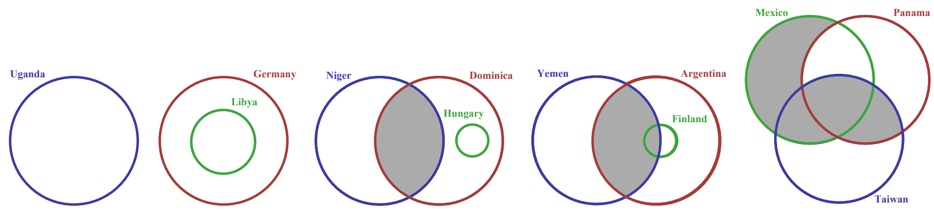

Fig. 8. The four Euler diagrams used for disjointness tasks.

Table 1. Answer positions for each question.

	Subset				Disjoint			
Treatment	WM	WMA	NWMA	NWM	WM	WMA	NWMA	NWM
Answer Position	1	4	2	3	4	1	3	2
Answer	$Z \subseteq E$	$Z \subseteq O$	$V \subseteq P$	$T \subseteq C$	$L \cap U = \emptyset$	$H \cap N = \emptyset$	$F \cap Y = \emptyset$	$M \cap T = \emptyset$

diagrams in the main study, except that the labelling differed. The diagrams were presented in a random order, generated for each participant, to reduce any potential ordering effect. Equal rankings were permitted and participants were asked to explain their ranking. The subset question included an attention check, with participants being asked to choose a specified option from a dropdown list.
Statistical Methods We collected accuracy and time data as indicators of performance, with accuracy viewed as more important than time: one treatment is judged to be more effective than another if users can perform tasks significantly more accurately with it or, if no significant accuracy difference exists, performance is significantly quicker when correct answers are provided. We employed a generalized estimating equations model [12] to analyse the accuracy and time data. For the preference analysis, we analysed data that related to the most preferred treatment. A local odds ratios GEE model [16] estimated the probability of each treatment being most preferred. The treatments were then compared pairwise, using the ratio of their associated probabilities. For the accuracy, time and preference data, it was not appropriate to apply commonly used parametric or non-parametric statistical method (e.g. ANOVA and Kruskal–Walis tests) because the data violated the normality assumption and the responses for each individual are expected to be correlated, and so not independent. The models and statistical output can be found at www.eulerdiagrams.com/wellmatched.

3.3 Euler Diagram Results

We report on the results of our evaluation of well-matchedness in Euler diagrams; the two studies (the other on linear diagrams) were run in parallel, with Prolific Academic participants being randomly exposed to either Euler or linear diagrams. We set pre-screening criteria: the first language had to be English, they had to have an approval rating of at least 98%, and have completed at least five studies on the Prolific platform. In addition, we only permitted the study to be taken on a desktop device, excluding the use of mobiles and tablets.

Each participant was paid £2.06 and told that we expected the study to take 15 minutes, with a maximum time allowed of 56 minutes (set by Prolific).

The pilot revealed that some questions had unexpectedly low accuracy rates. This led us to improve the training material at the beginning of the study, with additional explanation on the meaning of shading and new pages explaining the task answers. We also assigned a letter to each diagram in the preference phase and asked participants to use these letters when making comments, so that we could more accurately match their remarks to the diagrams. Lastly, we rectified an incorrect positioning of the correct answer to question 8. When gathering data for a second pilot, there was a technical issue, resulting in partial data being collected. Therefore, we ran a third pilot which still revealed low accuracy rates for some questions. Having already added material to the training given, it was felt that these low accuracy rates could be a feature of the treatments being evaluated, so we proceeded with the main data collection. The pre-screening criteria were carried forward with the additional criterion that no pilot participant could take part. As is standard, no participant could take part more than once.

We recruited 126 participants with the following distribution: 101 successfully completed, 0 were inattentive, and 25 failed to complete the study. Of the 101 participants who completed, 70 identified as female and 31 as male. Ages ranged from 18 to 69, with a mean of 34. Results are declared significant if $p \leq 0.05$. Note that we do not apply Bonferroni corrections. Some researchers routinely do so but corrections should only be applied when certain conditions are met [3][6].

Accuracy Analysis The mean accuracy rate overall was 61.39% with the treatment rates being: 86.63% for WM, 60.89% for WMA, 45.54% for NWMA, and 52.48% for NWM. These rates are indicative of performance differences but we must be mindful that the statistical methods employed do not compare them. When conducting our analysis, we found that there was no significant interaction between the treatment and the task type ($p = 0.174$), so we report on an analysis excluding the associated interaction term from the model. From this we derived the following ranking of treatments:

$$\textit{accuracy ranking: } WM > WMA > NWM > NWMA.$$

This matches our hypothesised ranking. For space reasons, we omit the p-values, which ranged from 0.0499 to < 0.0001.

Time Analysis The mean time taken overall was 30.34 seconds with the treatment means being: 21.79s for WM, 30.03s for WMA, 34.28s for NWMA, and 35.24s for NWM. For correct answers only, the overall mean was 26.16s, with the treatment means being: 20.29s for WM, 26.42s for WMA, 30.94s for NWMA, and 31.39s for NWM. Again, these rates are indicative of performance differences

[6] The goal of [3] is to provide advice, to researchers whose studies involve multiple testing, on when to use corrections: "[Bonferroni corrections] should not be used routinely and should be considered if: (1) a single test of the 'universal null hypothesis' (Ho) that all tests are not significant is required, (2) it is imperative to avoid a type I error, and (3) a large number of tests are carried out without preplanned hypotheses." None of these considerations apply in our case.

but the statistical methods employed do not compare them. When conducting our analysis, there was a significant interaction between the treatment and the task type ($p = 0.0108$), so we report on an analysis broken down task type:

time ranking for subset: WM > WMA = NWMA = NWM.
time ranking for disjoint: WM = WMA > NWMA = NWM.

This ranking is the partially consistent with our hypothesised ranking. In the significant cases, the subset p-values ranged from 0.0001 to < 0.0001 and for the disjoint analysis all were less than 0.0001.

Preference Analysis From the data provided by participants, we found an overwhelming preference for well-matched Euler diagrams, which were top-ranked 190 times. The other treatments were ranked top as follows: 17 times for WMA, 3 times for NWMA, and 11 times for NWM; recall joint rankings were permitted. When fitting our statistical model, we found that preference did not depend on task type ($p = 0.4715$) and, so, our results are based on a simplified model from which we obtained the following ranking:

preference ranking: WM > WMA = NWMA = NWM.

For space reasons, we omit the associated p-values, with those below the 5% threshold ranging from 0.0215 to < 0.0001. Comments made by participants often indicated that shading was confusing and highlighted their perceived simplicity of the diagrams that used only spatial relations between circles. Generally, the participants' comments supported the exploitation of spatial relationships between circles over shading.

Discussion For Euler diagrams, we can answer RQ1 and RQ2 affirmatively. Given that we view accuracy as the most important indicator of relative performance, our data also support [H1] to [H3]. In summary, the well-matchedness of meaning carriers does significantly impact task performance. It is particularly interesting that, when tasks required information to be obtained from non-well-matched meaning carriers, the presence of well-matched meaning carriers led to significantly worse accuracy performance than when there were no well-matched meaning carriers at all.

4 Evaluating Well-Matchedness in Linear Diagrams

The design of the linear diagrams study matched that of the Euler diagram study. The only difference was due to the notation. The linear diagrams used in the performance phase are equivalent to those in Figs 3 and 4, relabelled similarly to the Euler diagram study (see Figs 1 and 7 as well as Figs 2 and 8). The diagrams in Figs 3 and 4 were used in the preference phase.

4.1 Hypotheses

We immediately carry forward [H1] to the linear diagram case, since it is regarded that well-matched meaning carriers are more effective. We must further explore linear diagrams when considering [H2] and [H3] since lines do not intersect as circles do, a core feature of our earlier deliberations. We again focus on the extraction of the information that *everyone who visited Ukraine also visited Romania* and that *no one visited both India and Zambia*, this time from Figs 3 and 4. The observation that colour hue is favoured over shape (in this case, lines) holds for linear diagrams. However, whilst the lines never intersect each other they sometimes have line breaks, where more than one line segment represents a set. The Gestalt principle of similarity tells us that people will group together visual objects that share characteristics seeing them as 'belonging together'. This suggests that using varying colours for the lines could outweigh potential performance degradation arising from line breaks. Current empirical research into the impact of the number of line segments in task performance is, however, inconsistent [1, 14] and requires further investigation. Thus, there is no clear evidence that a hypothesised ranking of diagrams should be based on the presence of line breaks, which is a feature not directly related to well-matchedness.

Therefore, we focus our attention on the use of shading. As with Euler diagrams, the use of one colour for shading compared to varying hues for the lines renders the well-matched meaning carriers more salient than the non-well-matched meaning carriers. Further, the same reasoning as in the Euler diagrams' case can be applied to the use of non-well-matched meaning carriers to convey the required information in a diagram that also uses well-matched meaning carriers. Hence, we also carry forward [H2] and [H3] to the linear diagram case.

4.2 Linear Diagram Results

Regarding the pilots for the linear diagram study, they ran concurrently with the Euler diagram study. The adaptations and errors identified were reported in section 3. For the main study on linear diagrams, we recruited a total of 146 participants with the following distribution: 104 successfully completed, 3 were classified as inattentive, and 39 failed to complete. Of the 104 participants, 69 identified as female and 35 as male. Ages ranged from 18 to 67 (mean: 33).
Accuracy Analysis The accuracy rate overall was 63.22% with treatment rates: 84.13% for WM, 64.42% for WMA, 50.48% for NWMA, and 53.85% for NWM. When conducting our analysis, we found that there was no significant interaction between the treatment and the task type ($p = 0.5921$), so we report on an analysis excluding the interaction term from the model. We derived the following:

$$\textit{accuracy ranking: WM} > \textit{WMA} > \textit{NWMA} = \textit{NWM} \;.$$

This ranking is the same as our hypothesised ranking, except that NWMA and NWM are not distinguished. The p-values below the 5% threshold ranged from 0.0180 to < 0.0001.

Time Analysis The mean time taken overall was 27.79s with the treatment means being: 21.30s for WM, 27.47s for WMA, 32.52s for NWMA, and 29.87s for NWM. For correct answers only, the overall mean was 25.64s, with the treatments means being: 19.72s for WM, 27.00s for WMA, 32.00s for NWMA, and 27.31s for NWM. When conducting our analysis, there was a significant interaction between the treatment and the task type ($p < 0.0001$), so we report on an analysis broken down task type:

time ranking for subset: WM > WMA = NWMA = NWM.
time ranking for disjoint: WM > WMA = NWMA = NWM.

This ranking is the partially consistent with our hypothesised ranking. In the significant subset cases, the below-threshold p-values ranged from 0.0002 to $<$ 0.0001 and for the disjoint cases they were between 0.0315 and 0.0002.

Preference Analysis From the data provided by participants, we found an overwhelming preference for well-matched linear diagrams, which were top-ranked 187 times. The other treatments were ranked top as follows: 11 times for WMA, 7 times for NWMA, and 7 times for NWM. As with Euler diagrams, we found that preference did not depend on task type ($p = 0.0631$), so our results are based on a simplified model from which we obtained the following ranking:

preference ranking: WM > WMA = NWMA = NWM.

For space reasons, we omit the associated p-values, with those below the 5% threshold all < 0.0001.

Participants' comments again indicated that shading was confusing. In addition, comments alluded to the clarity of diagrams when spatial relationships between lines were used. A minor theme through the comments centred on line breaks, with some participants feeling that broken lines were problematic (when multiple line segments are used to represent a set). These comments are consistent with prior work, which suggests people perceive linear diagrams with more line segments as being more cluttered [2].

Discussion For linear diagrams, we also answer RQ1 and RQ2 affirmatively. However, our data supported [H1] and [H2] but not [H3]. We speculate about why this is different to the Euler diagram case. In linear diagrams, when extracting information about a number of sets, it is always possible to ignore the lines that represent any other sets. This is because the lines are laid out in parallel, with their relative x-coordinates conveying semantics, and they never intersect each other. In our study, only two lines and any present shading needed to be considered to correctly perform the task. In both the NWMA and NWM cases, the two lines involved in the task are in a non-well-matched relationship. This leads us to speculate that an irrelevant third set is not a distraction in linear diagrams. Hence, there is no distinguishing feature – from the perspective of well-matchedness – between the NWMA and NWM cases, providing a plausible reason as to why [H3] is not supported. Furthermore, this reasoning does not contradict [H1] or [H2]. In the case of [H1], we are comparing WM and WMA with NWMA and NWM: in the former two cases, the two lines are both well-matched and in the latter two they are both non-well-matched. For [H2], in each diagram the two relevant lines are well-matched, but in the WM case there is no distracting shading unlike the WMA case (e.g., in Fig. 4, part of the Zambia line occupies a shaded overlap in the WMA case but not in the WM case).

We contrast the discussion above with the Euler diagram case. In most Euler diagrams the circles intersect, which causes points of discontinuation, to form regions. To compare two sets in Euler diagrams remains straightforward if the two corresponding circles are in a (salient) well-matched relationship. However, the presence of the third (well-matched) circle in the NWMA case renders the task more difficult: the third circle is not easily ignored, unlike linear diagrams, due to the intersecting nature of the circles. Intersecting circles form an atomic component (single unit) unlike the separate lines in a linear diagram. Thus, we posit that the intersecting nature of the circles in Euler diagrams makes ignoring a third, irrelevant, circle, non-trivial. Hence, the discussions here may suggest reasons why [H3] was supported for Euler diagrams but not for linear diagrams.

5 Conclusion

A major goal of the Diagrams community is to better understand features of diagrams that make them effective. Through our examination of meaning carriers, we have exposed their potential importance in this context. By viewing meaning carriers as being well-matched, or otherwise, we have begun to explore the role of well-matchedness in Euler and linear diagrams. Our results suggest that extracting information from well-matched meaning carriers is significantly easier (as measured by accuracy) than in non-well-matched cases. A particularly striking result arose with Euler diagrams: when tasks required information to be obtained from non-well-matched meaning carriers, the presence of well-matched meaning carriers led to significantly worse accuracy performance than when there were no well-matched meaning carriers at all. By contrast, blending well-matched and non-well-matched meaning carriers in linear diagrams did not expose the same behaviour. This difference between notations is embodied in [H3] being supported for Euler diagrams but not for linear diagrams.

There are two key take-away messages: using only well-matched meaning carriers led to the best performance and using both well-matched and non-well-matched meaning carriers in a single diagram is sometimes problematic. In the latter case, it is necessary to consider how the syntax of the diagrams gives rise to meaning carriers and their role in information representation and potential ability to distract from the task at hand.

There is ample scope for further work. Specifically for Euler and linear diagrams, there is the potential for eye-tracking studies to either support or refute our speculation concerning why we had different results for [H3]. Evaluations are needed for a richer variety of meaning carriers and tasks and also to explore how participant familiarity with the notations impacts the results. Meaning carriers and well-matchedness should be further explored in other diagrammatic notations. One such example arises from the semantics assigned to Euler and linear diagrams. Extending their semantics so that regions have existential import means that regions represent non-empty sets. Under these semantics, new meaning carriers arise, such as the intersection between two circles representing a non-empty set. Understanding whether our results generalise are important for our continued exploration of the benefits of diagrammatic communication.

Acknowledgement Gem Stapleton is partially supported by EPSRC grant EP/T019603/1.

References

1. Alqadah, M.: The impact of clutter on the comprehension of set visualizations. PhD Thesis: University of Brighton (2017)
2. Alqadah, M., Stapleton, G., Howse, J., Chapman, P.: The perception of clutter in linear diagrams. In: Diagrams 2016. pp. 250–257. Springer (2016)
3. Armstrong, R.: When to use the Bonferroni correction. Ophthalmic Physiol Opt. **34**(5), 502–8 (2014)
4. Barwise, J., Etchemendy, J.: Visualization in mathematics, chap. Visual information and valid reasoning, pp. 8–23. Mathematical Association of America (1990)
5. Callaghan, C.: Interference and dominance in texture segregation: Hue, geometric form and line orienation. Perception & Psychophysics **4**(46), 299–311 (1989)
6. Chapman, P., Stapleton, G., Rodgers, P.: PaL diagrams: A linear diagram-based visual language. J. of Visual Languages and Computing **25**(6), 945–954 (2014)
7. Chapman, P., Stapleton, G., Rodgers, P., Michallef, L., Blake, A.: Visualizing sets: An empirical comparison of diagram types. In: Diagrams 2014, pp. 146–160. Springer (2014)
8. Chen, J., Menezes, N., Bradley, A., North, T.: Opportunities for crowdsourcing research on Amazon Mechanical Turk. Human Factors **5**(3) (2011)
9. Duncan, J., Humphreys, G.: Visual search and stimulus similarity. Psychological review **96**(3), 433 (1989)
10. Gurr, C.: Aligning syntax and semantics in formalisations of visual languages. In: IEEE Sym. on Human-Centric Computing Languages and Environments. pp. 60–61. IEEE (2001)
11. Harrower, M., C.Brewer: ColorBrewer.org: An Online Tool for Selecting Colour Schemes for Maps. The Cartographic J. **40**(1), 27–37 (2003), accessed November 2013
12. Liang, K.Y., Zeger, S.L.: Longitudinal data analysis using generalized linear models. Biometrika **73**, 13–22 (1986)
13. Peirce., C.: Collected Papers, vol. 4. Harvard University Press (1933)
14. Rodgers, P., Stapleton, G., Chapman, P.: Visualizing sets with linear diagrams. ACM Trans. on Computer-Human Interaction **22**(6), article 27 (28 pages) (2015)
15. Stapleton, G., Jamnik, M., Shimojima, A.: What makes an effective representation of information: A formal account of observational advantages. J. of Logic, Language and Information **26**(2), 143–177 (2017)
16. Touloumis, A., Agresti, A., Kateri., M.: Generalized estimating equations for multinomial responses using a local odds ratios parameterization. Biometrics **69**(3), 633–640 (2013)
17. Wagemans, J., Elder, J., Kubovy, M., Palmer, S., Peterson, M., Singh, M.: A century of gestalt psychology in visual perception: I. perceptual grouping and figure-ground organisation. Psychol Bull. **138**(6), 1172–1217 (2012)