Gem Stapleton
Andrew Blake
Jim Burton
Anestis Touloumis

Presence and Absence of Individuals in Diagrammatic Logics: An Empirical Comparison

Abstract. The development of diagrammatic logics is strongly motivated by the desire to make formal reasoning accessible to broad audiences. One major research problem, for which surprisingly little progress has been made, is to understand how to choose between semantically equivalent diagrams from the perspective of human cognition. The particular focus of this paper is on choosing between diagrams that represent either the *presence* or *absence* of individuals. To understand how to best make this choice, we conducted an empirical study. We found that representing the presence of individuals supported task performance either significantly better than, or no worse than, representing the absence of individuals. The particularly striking feature of our results was that representing the absence of individuals in a way that makes the diagram highly cluttered is detrimental to human cognition. As a result, diagrams with this feature should be avoided, but diagrams using presence (irrespective of diagram clutter) or low-cluttered absence can be used to support cognition in the context of the tasks performed in our study.

Keywords: Individuals, Presence, Absence, Clutter, Cognition, Diagrammatic logics.

1. Introduction

The study of diagrammatic logics has been prominent since Shin's work on Venn-I and Venn-II [33]. Other diagrammatic logics have since been developed with much of the related research being on their formal properties, including expressiveness, soundness and completeness. These logics include Euler diagrams [14,25,41], spider diagrams [17,36], Euler/Venn diagrams [39] and concept diagrams [18], as well as existential graphs [9,34].

A major research problem faced by the diagrams community is to understand how to choose between semantically equivalent, yet syntactically different, diagrams from the perspective of human cognition. Surprisingly little progress has been made, in contrast to the significant advances on the theoretical aspects of diagrammatic logics. Without a thorough understanding of how different choices of diagram impact cognition, it will not

Presented by **Jacek Malinowski**; *Received* March 23, 2016

Studia Logica (2017) 105: 787–815
DOI: 10.1007/s11225-017-9711-6

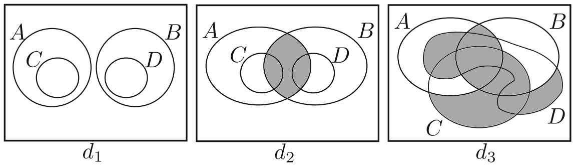

Figure 1. Visual clutter in Euler and Venn diagrams

be possible to fully exploit the established cognitive advantages of diagrams over symbolic and sentential notations [16,31,32]. One of the most prominent reasons for developing diagrammatic logics is to enable people to better understand information which provides further motivation for understanding the relative cognitive benefits of competing choices of diagrams.

A natural place to begin to understand syntactic choices is monadic first-order systems. Here, we need to understand how to best represent sets (via monadic predicates) and properties of sets such as the individuals they contain. The majority of diagrammatic logics exploit Euler diagrams to represent sets: each set is represented by a closed curve; the spatial relationships between the curves correspond to relationships between the sets. Further, they often employ syntactic devices, specifically labelled trees, to represent the presence of individuals: the region in which a tree is drawn indicates the set to which an individual belongs.

Recent years have seen the application of cognitive science and empirical methods to develop models that aim to explain task performance when using symbolic logics (for instance, [40]). Similar developments have taken place in the diagrams community, where empirical research has investigated how the choice of Euler diagram impacts cognition. Cluttered Euler diagrams [20] significantly reduce task performance [2]. Moreover, the use of shading, which typically denotes the emptiness of a set, can be detrimental when performing tasks [5,32]. For example, Figure 1 shows three semantically equivalent Euler diagrams, one of which is also a Venn diagram. The results of [2,5,32] imply that d_1 is sometimes the most effective representation from the perspective of cognition, and that the Venn diagram, d_3, can significantly hinder cognition compared to d_1 and d_2. Given this, we have insight into how to represent information about sets using Euler diagrams.

This paper takes the natural next step by investigating how we should represent individuals to aid cognition in logics based on Euler diagrams. Logics such as those in [6,12,18,21,37–39] all incorporate the representation of individuals using trees, so our investigation serves to underpin the use of constants in a variety of different logics. Figure 2 illustrates the presence of an individual: d_1 asserts that the individual a is in the set $A \cup B \cup C$, since

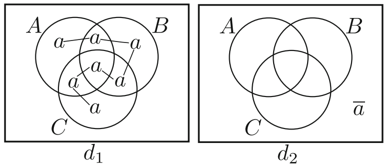

Figure 2. Presence versus absence

the tree whose nodes are written as 'a' is inside the region which represents $A \cup B \cup C$ (i.e. $A(a) \vee B(a) \vee C(a)$). Alongside representing the presence of individuals using labelled trees, Choudhury and Chakraborty introduced notation to assert the *absence* of an individual, a, from any particular set [6]. In Figure 2, d_2 represents the absence of a from $\overline{A \cup B \cup C}$, using $\overline{a}$, so d_2 directly expresses $\neg(\neg A(a) \wedge \neg B(a) \wedge \neg C(a))$, which is equivalent to $A(a) \vee B(a) \vee C(a)$. Thus, d_1 and d_2 are semantically equivalent.

A particular insight, as a consequence of the dual roles of presence and absence under a classical semantics, is that it is possible to exploit them both to reduce visual clutter; Choudhury and Chakraborty have also explored non-classical semantics for diagrams augmented with individuals [7,8], which is not the focus of this paper. For instance, d_1 in Figure 2 appears cluttered: there are many occurrences of a connected by lines. By contrast, using $\overline{a}$ in d_2 has allowed us to reduce visual clutter. As diagram clutter has been shown to hinder cognition in other contexts, it is therefore important to examine the interplay between clutter and the use of presence and absence. We do just that in this paper, which sets out to understand how to choose between representing individuals using presence and absence, focusing on the relative levels of clutter arising from the two choices.

Section 2 introduces some terminology and the notion of a clutter score. We make hypotheses about which representational choice is most effective in Section 3, in the context of tasks which involve reading diagrams. The design of our empirical study is described in Section 4 and its execution is detailed in Section 5. The statistical methods adopted to analyze our data are given in Section 6. We analyze the data in Section 7, where we also discuss threats to validity, and interpret the results in Section 8. The study materials and the data collected can be found on our website [35].

2. Syntax, Semantics and Clutter

The notation we evaluated augments Euler diagrams with syntax to represent the presence and absence of individuals, formalized in [4]. In Figure 3,

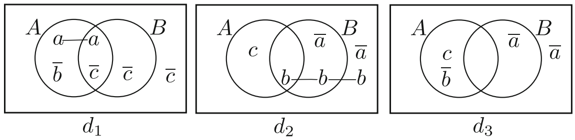

Figure 3. Presence and absence: i-sequences and $\bar{i}$-sequences

the diagram d_1 has two closed curves, labelled A and B, and makes assertions about three individuals. There is one i-sequence (i for individual), namely a, comprising two nodes joined by one edge; we say that an individual's *presence* is visualized if it is represented by an i-sequence. There are four $\bar{i}$-sequences, namely $\overline{b}$ and three $\overline{c}$s, each of which comprise a single node; we say that an individual's *absence* is visualized if it is represented by an $\bar{i}$-sequence. Consistent with [4,6], $\bar{i}$-sequences can only be of length 1.

The semantics are given via a standard model-theoretic approach [4]. In brief, an i-sequence asserts that the represented individual is in the set represented by the region that contains the i-sequence. Similarly, an $\bar{i}$-sequence asserts that the represented individual is *not* in the set represented by the containing region. The semantics are classical, so if an individual, a, is not in the set A, then $a \in \overline{A}$. Thus, d_1 in Figure 3 directly asserts that $a \in A$, $b \notin A\backslash B$, $c \notin A \cap B$, $c \notin B\backslash A$ and $c \notin \overline{A \cup B}$. This is equivalent to $a \notin \overline{A}$, $b \in \overline{A\backslash B}$, and $c \in A\backslash B$, directly expressed by d_2.

We can choose which *individual type* (i.e. an i-sequence or a set of $\bar{i}$-sequences) to use, affecting the clutter level in diagrams. Prior work devised a measure of clutter arising from individuals [4]: each i-sequence, $a - a \ldots a - a$, contributes n to the *clutter score* if the number of a symbols plus the number of connecting lines is n; each $\bar{i}$-sequence contributes 1 to the clutter score. In Figure 3, the clutter score for d_1 is 7 (3 from a and 1 from each of $\overline{b}$ and the three $\overline{c}$s). The diagram d_2 is derived from d_1 by 'swapping' the i-sequence for $\bar{i}$-sequences, and the $\bar{i}$-sequences for i-sequences. This has changed the clutter score to 8. The diagram d_3 is also semantically equivalent to d_1, yet is *minimally* cluttered, with a clutter score of 4.

3. Tasks and Hypotheses

We took the standard approach of collecting performance data (accuracy and time) from participants as a measure of cognitive effectiveness. The participants were presented with a set of diagrams. They were to answer one multiple choice question for each diagram, with each question beginning with the text *which one of the following statements is true*? The same three choices were always presented (paraphrased here):

Choice 1: The individual a is in the set A.

Choice 2: The individual a is not in the set A.

Choice 3: We do not know whether the individual a is in the set A.

The way in which the answers were presented is shown in Section 4.1 (Figure 7), and illustrated in the examples below.

Using this style of question, we wanted to establish whether diagram clutter or the use of presence or absence has a significant impact on task performance. We expected both of these properties to affect task performance and we also we expected the answer type to play an important role.

In order to provide a basis for our hypotheses, we appealed to the idea of *well-matchedness*, a concept introduced by Gurr [13]. Informally, a notation is well-matched if its syntactic relations mirror, in a homomorphic way, the semantic relations. Euler diagrams are an excellent example of a well-matched notation. For instance, curve containment mirrors set containment and curve (interior) disjointness mirrors set disjointness. Well-matchedness is considered to be a feature of diagrams that makes them preferable representations of information to traditional symbolic notations. In our work, we were particularly interested in the use of individuals to represent set membership of given elements. Like Euler diagram well-matchedness, the containment of an i-sequence (i.e. a presence individual) in a region directly mirrors the containment of the corresponding element in the represented set.

However, it may seem that the $\bar{i}$-sequence notation for absence is inherently counter-intuitive, and therefore possibly not well-matched, since the placement (i.e. existence) of a piece of syntax (i.e. the $\bar{i}$-sequence) in a region signifies that some element is not in the represented set. The notation for absence is, though, well-matched by Gurr's definition: we can establish a homomorphism between the concrete syntax and the direct semantics. Despite this, we expect crucial epistemological differences between presence and absence syntax. Using different terminology, Moktefi argues that well-matchedness is a necessary but not sufficient condition for a notation to support reasoning in "natural" ways (as Euler diagrams arguably do) [26]. In light of this, we treat well-matchedness as a continuum, rather than a binary property. We will appeal to well-matchedness in the context of individuals to derive hypotheses concerning the three answer choices.

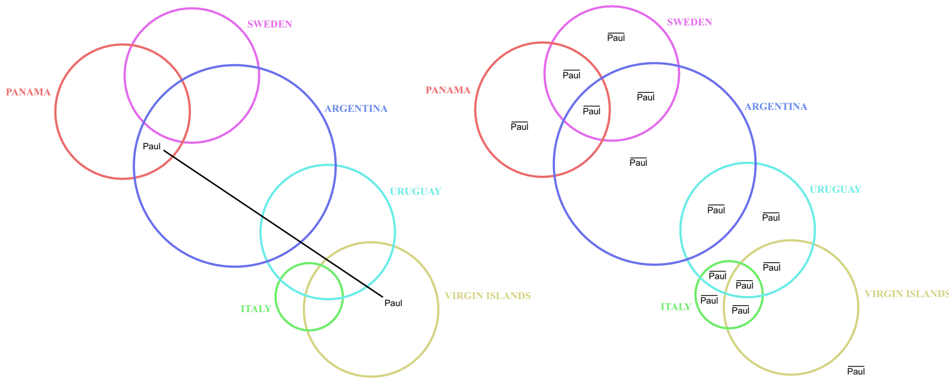

Figure 4. Contrasting low clutter (presence) and high clutter (absence)

A second basis for our hypotheses comes from research on pre-attentive processing [15] and visual search [10, 19, 30, 42]. This research suggests why it may be faster and less prone to error to identify the location of Paul in the left-hand diagram of Figure 4 compared to the right. For instance, Huang et al.'s Boolean Map theory [19] divides visual search into two phases, *selection* and *access*. In the left-hand diagram, the selection and access phase takes advantage of the low-level, pre-attentive visual system, since the tree labelled Paul is salient and a unique target. By contrast, using $\overline{\mathsf{Paul}}$ requires multiple syntactic elements to express the same information. When using the diagram on the right, the viewer is actually searching for regions in which $\overline{\mathsf{Paul}}$ does not appear, if they want to determine the set that contains Paul. The multiple occurrences of $\overline{\mathsf{Paul}}$ potentially inhibit the search task: the amount of clutter could increase the time taken by viewers in a visual search for regions not containing $\overline{\mathsf{Paul}}$. It should not be taken, however, that clutter only arises through the use of absence. Whilst a presence individual can be viewed as one syntactic item, it comprises nodes (the names) and edges, each of which are syntactic items in their own right. Therefore, both individual types can give rise to high levels of clutter in a diagram.

The visual search literature provides insight as to why clutter has a detrimental impact on task performance. Mechanisms and strategies of visual search utilise knowledge about targets (i.e. individuals) and their locations (i.e. the regions). Rosenholtz et al.'s work, for example considering excessive and disorganised display items [30], tells us that visual clutter impacts these strategies and is therefore detrimental to task performance during visual search. When a presence individual is the target, the larger the size of the region in which it is located, the more excessive the amount of syntax.

Regarding (dis)organisation, viewing the tree as a single entity leads to an organised display: one can visually follow the connected nodes via the edges.

By contrast, absence individuals are disconnected and, thus, disorganised leading to potential increased difficultly when searching. This difficulty is intertwined with the number of absence individuals: few occurrences and, thus, low clutter leads to more organisation whereas excessive occurrences and, thus, high clutter, leads to more disorganisation. In summary, presence individuals are organised whereas absence individuals exhibit degrees of disorganisation but both individual types can give rise to varying clutter levels. We hold the overarching view that when a diagram has a lower level of visual clutter than another it can more effectively support task performance.

3.1. Hypotheses for Answer Choice 1

To begin this section, we will expand on how individual types are well-matched to their semantics. Presence individuals explicitly represent a given element and are located in the region corresponding to the set that contains it. For example, in Figure 4, the *presence diagram* (on the left) explicitly represents the individual Paul. The location of Paul inside two zones expresses that Paul is interested in either Argentina and Panama only *or* interested in the Virgin Islands only. This containment of Paul in these two zones directly mirrors the containment of Paul in the corresponding set. In this sense, the diagram is well-matched when depicting element containment in sets.

Furthermore, the transitive property of syntactic inclusion mirrors the transitive property of set membership: if $x \in A$ and $A \subseteq B$ then $x \in B$. Diagrammatically, referring again to Figure 4, we can see that Paul is also included in the region which represents the set Argentina ∪ Panama ∪ Virgin Islands; this corresponds to the diagram explicitly representing answer choice 1. That is, if the question 'which of the following statements are true?' is asked of the presence diagram in Figure 4, with options

1. Paul is interested in Argentina, Panama, or the Virgin Islands
2. Paul is not interested in Argentina, Panama, or the Virgin Islands
3. Do not know whether Paul is interested in either Argentina, Panama, or the Virgin Islands

then the diagram explicitly represents the first choice, Paul is interested in Argentina, Panama, or the Virgin Islands, in a well-matched way.

In the case of absence individuals, they also explicitly represent the given element and are, between them, located in the region corresponding to the set that does *not* contain this element. For example, in Figure 4, the absence diagram (on the right) explicitly represents the individual Paul. Each zone that contains $\overline{\text{Paul}}$ expresses that Paul is not an element of the corresponding set. For instance, the location of $\overline{\text{Paul}}$ in the zone inside both, and only, Panama and Sweden asserts, *directly*, that Paul is not interested in Panama and Sweden but nothing else. There is clearly a case to be made that the use of absence is not well-matched when we wish to identify the set that contains Paul: the region that represents this set contains no representation of Paul at all. Therefore, regarding answer choice 1, we make the following two hypotheses:

H1: *Low clutter presence diagrams support significantly better task performance than high clutter absence diagrams.* The basis for this hypothesis is that low clutter presence diagrams are more effective because: (a) they are well-matched to answer choice 1 whereas high clutter absence diagrams are not, and (b) they are low in clutter unlike high clutter absence diagrams.

H2: *There is no significant difference in task performance between high clutter presence diagrams and low clutter absence diagrams.* The basis for this hypothesis is that: (a) high clutter presence diagrams could be more effective because they are well-matched to answer choice 1 whereas high clutter absence diagrams are not, and (b) low clutter absence diagrams could be more effective because they are low in clutter unlike high clutter presence diagrams. Thus, there is no reason to suppose that one class of diagram is the most effective representation in this case.

As seen in H2, it is not clear whether well-matchedness or visual clutter has more influence over relative task performance. If H2 is not supported by our study, it may help to shed light on the relative trade-off between clutter level and well-matchedness for diagrams of this type.

3.2. Hypotheses for Answer Choice 2

Focusing now on answer choice 2, which is phrased as 'the individual a is not in the set A', it can again be argued that presence diagrams are well-matched. For example, the *exclusion* of the individual Amy, in the presence diagram in Figure 5, from eight of the zones asserts that Amy is *not interested* in the corresponding combinations countries. For instance, the fact that Amy is not placed in curve labelled Turkey expresses that Amy is not in the set Turkey. Well-matchedness arises because the region that does not contain Amy represents a set that does not contain Amy: the diagram *directly*

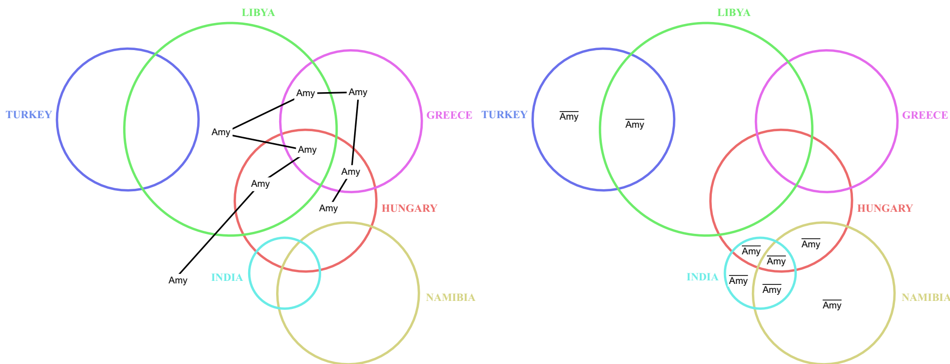

Figure 5. Contrasting presence and absence for choice 2 answers

expresses the absence of Amy from this set because of the absence of Amy from the corresponding region.

For absence diagrams, however, a case can be made for well-matchedness when we want to know the set in which an individual does not lie. In particular, to determine whether $x \notin A$, one can locate the set A and 'see' whether $\overline{x}$ is located there. For example, in Figure 5, the absence diagram *directly* expresses the semantics 'Amy is not interested in India, Namibia or Turkey'. In this sense, the syntax is well-matched to the intended semantic interpretation. However, things are not this clear cut. In this case, the transitivity of syntactic inclusion does not transfer across to the semantic level: $x \notin A$ (diagrammatically, $\overline{x}$ is inside A) and $A \subseteq B$ (diagrammatically, A is enclosed by B) does not imply $x \notin B$. Therefore, we could view this use of absence, when the information represented corresponds to answer choice 2, as in some sense less well-matched than the use of presence.

In making our hypotheses, we recognise that participants will have been trained in the interpretation of these diagrams. Moreover, the meaning of the absence diagrams directly represents choice 2. Therefore, we make the following two hypotheses, although whether they are likely to be supported is perhaps less likely than for H1 and H2:

H3: *Low clutter presence diagrams support significantly better task performance than high clutter absence diagrams.* The basis for this hypothesis is that: (a) both representations are arguably well-matched to this answer choice, but (b) low clutter presence diagrams are, obviously, low in clutter unlike high clutter absence diagrams.

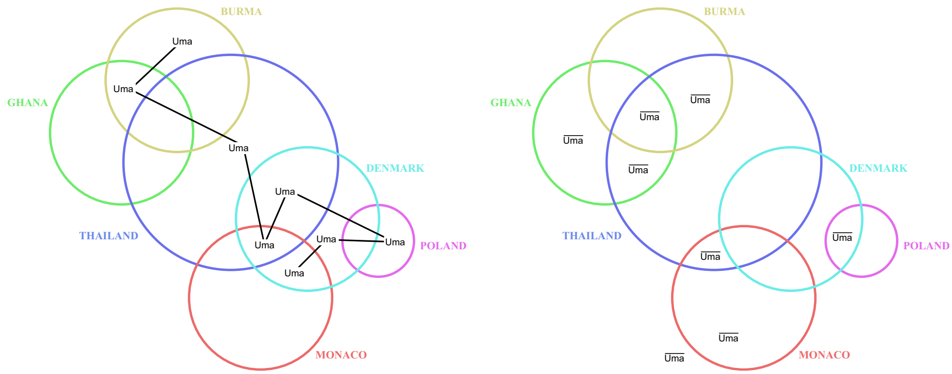

Figure 6. Contrasting presence and absence for choice 3 answers

H4: *Low clutter absence diagrams support significantly better task performance than high clutter presence diagrams.* The basis for this hypothesis is that: (a) both representations are arguably well-matched to this answer choice, but (b) low clutter absence diagrams are, obviously, low in clutter unlike high clutter presence diagrams.

Given the discussion above concerning levels of well-matchedness, it will be particularly interesting to see if these hypotheses hold.

3.3. Hypotheses for Answer Choice 3

Answer choice 3 is different to answer choices 1 and 2 in that it captures uncertainty: it is correct when the diagram does not express one of $x \in A$ and $x \notin A$. In the case of presence diagrams, the answer is choice 3 when part of the individual is in the region that represents the set A and part of it is outside this region, directly mirroring the uncertainty captured by choice 3. For example, in the presence diagram in Figure 6, part of Uma is inside the region that represents Burma ∪ Ghana ∪ Monacco and part of Uma is outside this region. This visually illustrates the uncertainty about whether Uma is interested in Burma ∪ Ghana ∪ Monacco in a well-matched way. In the absence diagram in Figure 6 there are zones in the region representing Burma∪Ghana∪Monacco that do not contain occurrences of $\overline{\text{Uma}}$—so Uma could inhabit the set represented by one of these zones. Similarly to answer choice 1, the use of absence is not well-matched here: the diagram gives no explicit indication that Uma could be in one of these sets.

Tying this discussion together, we make the following hypotheses:

H5: *Low clutter presence diagrams support significantly better task performance than high clutter absence diagrams.* The basis for this hypothesis is that low clutter presence diagrams are: (a) well-matched to answer choice 3, unlike high clutter absence diagrams, and (b) low in clutter unlike high clutter absence diagrams.

H6: *There is no significant difference in task performance between high clutter presence diagrams and low clutter absence diagrams.* The basis for this hypothesis is that (a) high clutter presence diagrams could be more effective as they are well-matched to answer choice 3, whereas low clutter absence diagrams are not, and (b) low clutter absence diagrams could be more effective because they are low in clutter unlike high clutter presence diagrams. Thus, there is no reason to suppose that one class of diagram is the most effective representation in this case.

4. Experiment Design

In this study, congruent with [1,24,27,28,32], we viewed comprehension in terms of task performance: one diagram is more comprehensible than another if users can interpret it significantly more accurately or, if no difference in accuracy exists, significantly more quickly. To gather accuracy and time data, participants provided answers to multiple choice questions. Each diagram contained information using just one individual type.

Initially, we adopted a mixed design with two participant groups. One group saw half of the diagrams containing presence information and half of those containing absence information. The other group saw the same Euler diagrams, but with the presence information swapped for absence information and, likewise, the absence information swapped for presence information. Participants were also exposed to both high and low cluttered diagrams. A pilot study (reported on later) had an error rate that was higher than expected and participants commented on the difficulty of understanding both presence and absence. Given these two insights, we modified the design: each participant saw both high and low cluttered diagrams, but was only exposed to either presence or absence, but not both.

4.1. Information Context

Previous empirical studies on the interpretation of logical diagrams [11] and diagrams used purely for information visualization [29] deemed it important to use a real-world scenario for the information being conveyed: the use of symbols can be off-putting to those without formal training in logic. We

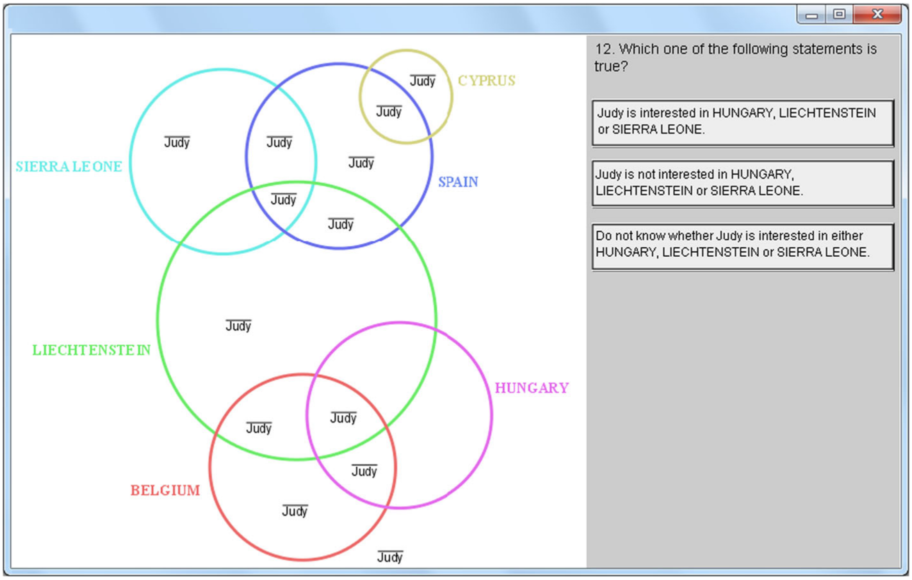

Figure 7. A screenshot showing how the question and diagram were displayed

adopted the same approach: our diagrams represented information about the countries in which people were interested. Moreover, it was important to avoid any possibility of previous knowledge of the data impacting the results, so all diagrams conveyed fictitious information from our real-world scenario. The way in which questions were displayed to participants can be seen in Figure 7, which is a screenshot of our data collection software. For the displayed diagram, the answer is choice 1.

4.2. Diagrams for the Study

As we were interested in the impact on comprehension of the choice between presence and absence, as well as diagram clutter, our study required a range of diagrams to be drawn. These diagrams needed to exhibit presence, absence and a variety of clutter levels; the diversity of our set of diagrams was deemed important for the generalizability of our results.

When drawing the diagrams, care was taken to ensure that their layouts aided cognition [3,13,29]. The following Euler diagram drawing guidelines were followed: each set was represented by a circle; each circle had a three pixel stroke width and a unique colour; their interiors had a transparent fill; the areas of the zones ensured that the individuals' names would comfortably fit in them; all set names were written in capital letters and took the same colour as their associated circle; the names were chosen so that no two names had a similar pronunciation. The last guideline was designed to reduce the potential for errors due to misreading; such errors could lead to incorrect answers that were not due to clutter or the use of presence or absence.

We also had to decide how many sets to visualize and how many zones to include in the diagrams used in the study. We considered the following:

We wanted to ensure that we could display both high and low clutter diagrams for both presence and absence. Given the clutter measure, if there are n zones in a diagram and an i-sequence with clutter score m then the set of $\bar{i}$-sequences, obtained by swapping the i-sequence, has a clutter score of $n - \frac{m-1}{2}$. To ensure we had a clear difference between m and $n - \frac{m-1}{2}$ (thus distinguishing between high and low clutter), a reasonable number of zones needed to be present. Given the topological constraints imposed when using circles, this meant a reasonable number of sets had to be visualized.

We wanted our questions to be non-trivial, so that cognitive effort was needed to answer them. This required more than one set to be involved in the multiple choice answers. However, if too many sets were involved, the tasks could become too difficult. This could result in high error rates or increase the variability in the time taken due to the complexity of the Euler diagrams, rather than being attributable to the evaluated diagrammatic syntax. It was important to reduce such unwanted variance in our data.

We wanted to ensure that the number of sets involved in the answer to our question did not give rise to a pattern that could indicate the correct answer. Such a pattern could give rise to a learning effect or the correct answer could be identified without the need for reading the diagram, potentially biasing or invalidating our results. Therefore, a fixed number of sets was involved in the answer to each question.

Taking the above considerations into account, all diagrams used in the data collection phase of the study visualized six sets, with three sets involved in every multiple choice answer, and had 16 zones.

Once the Euler diagrams had been drawn, we had to add i-sequences and $\bar{i}$-sequences to the diagrams. We decided that each diagram would only make a statement about a single individual, to isolate the effect of using presence and absence without extraneous diagrammatic elements potentially distracting from the task that was undertaken. We adopted the following conventions: individuals' names were placed close to the centre of their containing zone, so far as was possible; for i-sequences, the connecting lines had a two pixel stroke width; for $\bar{i}$-sequences, the overlines had a one pixel stroke width and ran the entire length of the name; the names (and lines) were coloured black, to clearly distinguish them from the circles, and were written lower

Table 1. Assignment of clutter scores and answer choices to diagrams

Diagram number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	18	20	21	22	23	24
Presence CS	**15**	**15**	**17**	**17**	3	3	5	5	**15**	**15**	**15**	**15**	3	3	5	5	**15**	**15**	**17**	**17**	3	3	5	5
Absence CS	8	8	7	7	**14**	**14**	**13**	**13**	8	8	8	8	**14**	**14**	**13**	**13**	8	8	7	7	**14**	**14**	**13**	**13**
Choice	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3

case, except for the first letter; the names were randomly generated and were culturally diverse.

Initially, we drew 24 Euler diagrams for the main data collection phase of the study. Each diagram was copied to create a further 24 diagrams. Each original diagram was assigned an i-sequence and its copy was assigned a set of $\bar{i}$-sequences such that the pair of diagrams were semantically equivalent. Table 1 shows the clutter scores arising from the i-sequences, thus representing presence, or the sets of $\bar{i}$-sequences, thus representing absence. High clutter scores ranged from 13 to 17, and low clutter scores ranged from 3 to 8; in Table 1, the high clutter scores are in bold. The 'choice' row indicates the correct answer to the question 'which one of the following statements is true'. We ensured an even distribution of high and low clutter scores arising from presence and absence across each answer choice.

5. Experiment Method

We ran a first pilot study with a mixed design, to which we recruited six participants (three per group). Recall that, with our initial design, each participant saw high and low clutter diagrams and both presence and absence of individuals. The pilot revealed a high error rate, 35 incorrect answers out of 144 responses (24.3%), leading us to redesign the study. Again, recall that after this redesign each participant saw both high and low cluttered diagrams and was exposed to either presence or absence, but not both. We ran a second pilot, recruiting four participants. This yielded 11 errors out of 96 responses (11.5%). Satisfied that there were no other issues with the study design, we proceeded with the main study, for which 60 participants (44 M, 16 F; ages 18–38, mean 22.5) were recruited. All participants were students from the University of Brighton; none reported a sight-based disability and none were members of the authors' research group.

The participants undertook the study in a usability laboratory which provided a quiet environment free from interruption. Bespoke software was written to gather performance data. The same computer and monitor were used by each participant. The monitor had a high resolution, ensuring that the colours used in the Euler diagrams were readily visible and distinguishable. Each participant was alone during the experiment, except for an experiment facilitator who was present throughout. Each participant was requested not to discuss the details of the study with other people after they had taken part. The participants were informed that they could withdraw at any time. Each participant completed the experiment in under 1 h.

The study had three main phases: paper-based training, software training, and the main data collection phase. In the paper-based training phase all participants were treated as having no previous experience of Euler diagrams with individuals and were given the same training. Participants were introduced to the notion of individuals in Euler diagrams using hard-copy printouts of three diagrams, none of which were used in the subsequent experiment phases. Those answering questions about diagrams representing the presence of individuals received training in that notation, but not absence. Similar training was given to participants in the absence group.

The second phase provided training on how to use the data collection software. Participants were shown three questions, one for each answer type, and asked to attempt them in the software. If a question was answered incorrectly, the facilitator explained the answer to the participant to increase their understanding. As with the paper-based training phase, these diagrams and questions were not reused during the third (final) study phase.

During the third phase, we collected performance (accuracy and time) data. The 24 questions were displayed in a random order. After choosing an answer, the software would move to a pause screen, asking the participant to click when they were ready to start the next question. If an answer was not provided within 2 min, the pause screen would be shown and a *timeout* was recorded; the time limit was set to ensure that the experiment ended within reasonable time. On completing the study, the participants were given a £6 canteen voucher to thank them for their time.

6. Method of Statistical Analysis

We employed a GEE based statistical model [23] that allowed us to estimate the odds of producing a correct answer with the different combinations of individual type, clutter level, and answer choice:

$$\begin{aligned}\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) &= \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \beta_3 x_{ij3} + \beta_4 x_{ij4} \\ &\quad + \beta_5 x_{ij1} x_{ij2} + \beta_6 x_{ij1} x_{ij3} + \beta_7 x_{ij1} x_{ij4} + \beta_8 x_{ij2} x_{ij3} \\ &\quad + \beta_9 x_{ij2} x_{ij4} + \beta_{10} x_{ij1} x_{ij2} x_{ij3} + \beta_{11} x_{ij1} x_{ij2} x_{ij4}\end{aligned}$$

where: π_{ij} is the probability for subject i ($i = 1, \ldots, 60$) to answer correctly question j ($j = 1, \ldots, 24$); x_{ij1} is the indicator that the diagram given to subject i for answering question j contained presence; x_{ij2} is the indicator that a low cluttered diagram was given to subject i for answering question j; x_{ij3} is the indicator that question j was of Choice 2; and x_{ij4} is the indicator that question j was of Choice 3. With this GEE based statistical model, we could determine whether the odds of providing a correct answer for one combination of individual type, clutter level and answer choice was significantly different from other combinations while taking into account the expected correlation among the responses provided by each individual participant. Statistical output is included in the supplementary material [35]; we report on the main findings in the following section.

We employed another GEE based statistical model for the time data in order to estimate the time taken to provide a correct answer with the different combinations of individual type, clutter level, and answer choice:

$$\begin{aligned}\log(Y_{ij}) = {} & \gamma_0 + \gamma_1 x_{ij1} + \gamma_2 x_{ij2} + \gamma_3 x_{ij3} + \gamma_4 x_{ij4} \\ & + \gamma_5 x_{ij1} x_{ij2} + \gamma_6 x_{ij1} x_{ij3} + \gamma_7 x_{ij1} x_{ij4} + \gamma_8 x_{ij2} x_{ij3} \\ & + \gamma_9 x_{ij2} x_{ij4} + \gamma_{10} x_{ij1} x_{ij2} x_{ij3} + \gamma_{11} x_{ij1} x_{ij2} x_{ij4}\end{aligned}$$

where: Y_{ij} is the time that subject i ($i = 1, \ldots, 60$) needs to answer question j ($j = 1, \ldots, 24$) correctly; and the covariates x_{ij1}, x_{ij2}, x_{ij3} and x_{ij4} are defined in the model for the accuracy data. In a similar manner as with the accuracy data, the GEE based statistical model for the time data allowed us to determine whether the time taken to provided a correct answer for one combination of individual type, clutter level and answer choice was significantly different from other combinations. Further details and statistical output are included in the supplementary material [35]; as with the accuracy analysis, we report on the main findings in the following section.

7. Results and Discussion

The results are based on data collected from 60 people, each answering 24 questions. For the accuracy analysis, we took the responses for which an answer was provided within the 2 min allowed, thus excluding only 2 (non-) responses. There were two timeouts, both for low clutter presence diagrams and answer choice 3; they arose from different participants. Of the remaining 1438 responses, there were a total of 398 errors giving an overall error rate of 27.7% and, therefore, accuracy rate of 72.3%. Although the overall error rate was found to be higher than the reduced error rate in the second pilot study (from 22.9% to 11.5%), this estimate is likely to be more precise as the main study involved more participants than in the pilot study. We analyzed only the time data for which a correct answer was provided, consistent with previous research such as [24]. When we determined which combination of treatments most effectively supported task performance, we viewed accuracy as a more important performance indicator than time. This meant that one combination of treatments was taken to be more effective than another if it was significantly more likely to yield a correct answer. Otherwise, we appealed to differences in the time taken to provide a correct answer; in any case, we present all time analysis for completeness. Throughout, we used a 5% significance level to call results statistically significant.

7.1. Results Concerning Hypothesis 1

Hypothesis 1 concerned *answer choice 1* and conjectured that *low clutter presence diagrams* support significantly better task performance than *high clutter absence diagrams*. Low clutter presence diagrams yielded 55 errors and 65 correct responses, giving an accuracy rate of 54.2%, and a mean response time of 23.2 s to provide a correct answer. High clutter absence diagrams yielded 86 errors and 34 correct answers, with an accuracy rate of 28.3% and a mean response time of 29.6 s.

Using the GEE based statistical model for the accuracy data, we estimated a 95% confidence interval (CI) for the odds of providing a correct answer with low clutter presence diagrams compared to high clutter absence diagrams, as well as a p value that allowed us to determine whether these two combinations of treatments were significantly different for answer choice 1. The estimated odds of correctly answering questions with low cluttered presence diagrams was 2.9893 times higher than that of high clutter absence diagrams with a 95% CI of $(1.3247, 6.7457)$ and p value of 0.0084. Therefore, low clutter presence diagrams supported significantly better task performance, in terms of accuracy, than high clutter absence diagrams.

Using the GEE based statistical model for the time data, we estimated a 95% CI for the ratio of the time (measured in seconds) needed to answer a question correctly with one combination of the treatments to that of another. The CI and its corresponding p value allowed us to determine whether these two combinations of treatments were significantly different for answer choice 1. The model estimated that the time needed to answer a question correctly with a low cluttered presence diagram was 0.7439 times that with a high clutter absence diagram with a 95% CI of $(0.5719, 0.9675)$ and p value of 0.0274. Therefore, low clutter presence diagrams supported significantly better task performance, in terms of time, than high clutter absence diagrams. The accuracy and time results both supported H1. We may suggest that *low clutter presence diagrams* allowed significantly better task performance than *high clutter absence diagrams* for *answer choice 1.*

7.2. Results Concerning Hypothesis 2

Hypothesis 2 concerned *answer choice 1* and conjectured that there was no significant difference in task performance between *high clutter presence diagrams* and *low clutter absence diagrams.* The accuracy rate for high clutter presence diagrams was 67.5% given 39 errors and 81 correct responses, and a mean response time of 22.5 s to provide a correct answer. Low clutter absence diagrams yielded 27 errors and 93 correct answers, with an accuracy rate of 77.5%, and a mean completion time of 20.6 s. The GEE based statistical model for the accuracy data implied that the estimated odds of correctly answering questions with high cluttered presence diagrams were 0.6030 times that of low clutter absence diagrams with a 95% CI of $(0.2359, 1.5412)$ and p value of 0.2907. Therefore, there was no significant difference between high clutter presence diagrams and low clutter absence diagrams for answer choice 1 with respect to accuracy. The GEE based statistical model for the time data implied that the time needed to answer a question correctly with a high cluttered presence diagram was 1.1139 times higher than that with a low clutter absence diagram but with a 95% CI of $(0.8903, 1.3935)$ and p value of 0.3454. Therefore, there was no statistically significant difference between high clutter presence diagrams and low clutter absence diagrams for answer choice 1 with respect to time. The accuracy and time results both support H2. We may suggest that there was no significant difference in task performance between *high clutter presence diagrams* and *low clutter absence diagrams* for *answer choice 1.*

7.3. Results Concerning Hypothesis 3

Hypothesis 3 concerned *answer choice 2* and conjectured that *low clutter presence diagrams* supported significantly better task performance than *high clutter absence diagrams.* Low clutter presence diagrams yielded 6 errors and 114 correct responses, giving an accuracy rate of 95.0%, and a mean response time of 15.9 s to provide a correct answer. High clutter absence diagrams yielded 11 errors and 109 correct answers, with an accuracy rate of 90.1% and a mean completion time of 20.8 s.

The GEE based statistical model for the accuracy data implied that the estimated odds of correctly answering questions with low cluttered presence diagrams compared to high clutter absence diagrams were 1.917 times higher but with a 95% CI of $(0.594, 6.192)$ and p value of 0.276. Therefore, there was no significant difference between high clutter presence diagrams and low clutter absence diagrams for answer choice 2 with respect to accuracy. The GEE based statistical model for the time data estimated that the time needed to answer correctly a question with a low cluttered presence diagram was 0.7932 times that needed with a high clutter absence diagram, with 95% CI of $(0.6437, 0.9775)$ and p value of 0.0297. Therefore, low clutter presence diagrams supported significantly better task performance, in terms of time, than high clutter absence diagrams. In this case, our secondary performance indicator—time—supports H3. We may suggest that *low clutter presence diagrams* allowed significantly better task performance than *high clutter absence diagrams* for *answer choice 2.*

7.4. Results Concerning Hypothesis 4

Hypothesis 4 focused on *answer choice 2* and conjectured that *low clutter absence diagrams* supported significantly better task performance than *high clutter presence diagrams*. Low clutter absence diagrams yielded 14 errors and 106 correct responses, giving an accuracy rate of 88.3%, and a mean response time of 18.3 s to provide a correct answer. High clutter presence diagrams yielded 12 errors and 108 correct answers, with an accuracy rate of 90.0%, and a mean completion time of 16.6 s. The GEE based statistical model for the accuracy data estimated that the odds of correctly answering questions with high cluttered presence diagrams compared to low clutter absence diagrams were 1.189 times higher but with a 95% CI of $(0.557, 2.538)$ and p value of 0.655. Therefore, there was no significant difference between high clutter presence diagrams and low clutter absence diagrams for answer choice 2 with respect to accuracy. The GEE based statistical model for the time data estimated that the time needed to answer correctly a question with a high cluttered presence diagram was 0.8687 times that with a low clutter absence, with a 95% CI of $(0.7082, 1.0655)$ and p value of 0.1766. Considering both the accuracy and time analysis, we have no evidence to support H4. There was no significant difference in task performance when using low clutter absence diagrams and high clutter presence diagrams when the answer is choice 2.

7.5. Results Concerning Hypothesis 5

Hypothesis 5 concerned *answer choice 3* and conjectured that *low clutter presence diagrams* supported significantly better task performance than *high* *clutter absence diagrams*. Low clutter presence diagrams yielded 25 errors and 93 correct responses, giving an accuracy rate of 78.8%, and a mean response time of 25.0 s to provide a correct answer. High clutter absence diagrams yielded 66 errors and 54 correct answers, with an accuracy rate of 45.9%, and a mean completion time of 27.8 s.

The GEE based statistical model for the accuracy data estimated that the odds of correctly answering questions with low cluttered presence diagrams compared to high clutter absence diagrams were 4.55 times higher with a 95% CI of $(2.22, 9.29)$ and p value of <0.001. Therefore, low clutter presence diagrams supported significantly better task performance, in terms of accuracy, than high clutter absence diagrams. The GEE based statistical model for the time data estimated that the time needed to answer correctly a question with a low cluttered presence diagram was 0.8549 times that with a high clutter absence diagram but with a 95% CI of $(0.6963, 1.0496)$ and p value of 0.1343. Therefore, there was no significant difference between low clutter presence diagrams and high clutter absence diagrams for answer choice 3 with respect to time. In this case, our primary performance indicator—accuracy—supported H5. We may suggest that *low clutter presence diagrams* allowed significantly better task performance than *high clutter absence diagrams* for *answer choice 3*.

7.6. Results Concerning Hypothesis 6

Hypothesis 6 concerned *answer choice 3* and conjectured that there was no significant difference in task performance between *high clutter presence diagrams* and *low clutter absence diagrams*. High clutter presence diagrams yielded 26 errors and 94 correct responses, giving an accuracy rate of 78.3%, and a mean completion time of 19.3 s to provide a correct answer. Low clutter absence diagrams yielded 31 errors and 89 correct answers, with an accuracy rate of 74.2%, and a mean completion time of 21.7 s.

The GEE based statistical model for the error data estimated that the odds of correctly answering questions with high cluttered presence diagrams compared to low clutter absence diagrams were 1.259 higher with a 95% CI of $(0.524, 3.027)$ and p value of 0.606. Therefore, there was no significant difference between high clutter presence diagrams and low clutter absence diagrams for answer choice 3 with respect to accuracy. The GEE based model for the time data estimated that the time needed to answer a question correctly with a high cluttered presence diagram compared to that with a low clutter absence diagrams was 0.8754 times with a 95% CI of $(0.7297, 1.0502)$ and a p value 0.1520. Therefore, there was no significant difference between high clutter presence diagrams and low clutter absence diagrams for answer choice 3 with respect to time. The accuracy and time results both support H6. We may suggest that there was no significant difference in task performance between *high clutter presence diagrams* and *low clutter absence diagrams* for *answer choice 3*.

7.7. Summary of the Results

Of the six hypotheses we made, drawing on the notions of well-matchedness relating to the use of individuals, clutter levels and visual search, the following five were supported by our empirical study:

H1: *Low clutter presence diagrams support significantly better task performance than high clutter absence diagrams when the answer is choice 1.*

H2: *There is no significant difference in task performance between high clutter presence diagrams and low clutter absence diagrams when the answer is choice 1.*

H3: *Low clutter presence diagrams support significantly better task performance than high clutter absence diagrams when the answer is choice 2.*

H5: *Low clutter presence diagrams support significantly better task performance than high clutter absence diagrams when the answer is choice 3.*

H6: *There is no significant difference in task performance between high clutter presence diagrams and low clutter absence diagrams when the answer is choice 3.*

There was no evidence, however, to support H4: *low clutter absence diagrams support significantly better task performance than high clutter presence diagrams when the answer is choice 2.* Here, there was no significant difference between the two combinations of individual type and clutter level.

The statistical analysis delivered two striking messages. Firstly, low clutter presence diagrams outperform high clutter absence diagrams, irrespective of answer choice. Secondly, low clutter absence diagrams did not perform significantly differently from high cluttered presence diagrams. These are the major findings of the study and lead to the following guidelines:

1. High clutter absence diagrams (overall accuracy rate: 54.7%; mean time 24.2 s) should be swapped for low clutter presence diagrams (overall accuracy rate: 76.0%; mean time 20.8 s). In this case, we saw a large effect

size for correct answers—with low clutter presence diagrams yielding 22 more correct answers for every 100 answers—and a modest effect size for time—with low clutter presence diagrams yielding, on average, a 3.4 s time saving.

2. Low clutter absence diagrams (overall accuracy rate: 80.0%; mean time 20.1 s) are equally effective as high clutter presence diagrams (overall accuracy rate: 78.6%; mean time 19.2 s). Therefore, either can be used without being detrimental to task performance.

7.8. Threats to Validity

Threats to validity are categorized as internal, construct and external [27]. Internal validity considers whether confounding factors, such as carry-over effect, affect the results. Construct validity considers whether the independent and dependent variables yield an accurate measure to test our hypotheses. External validity examines the extent to which we can generalize the results of the study.

With regard to internal validity the major consideration related to carry-over effect. In a repeated measure experiment this threat occurs when the measure of a treatment is affected by the previous treatment. To manage this effect, our final study design ensured that each participant was only exposed to either presence or absence, but not both: participants in one group saw all of the diagrams with clutter scores given in the 'Presence CS' row of Table 1; the other group saw those diagrams associated with the 'Absence CS' row. Therefore, there was no threat of carry-over effect from the presence and absence treatments. Regarding the clutter scores treatment, no participant was exposed to a high and low clutter pair of diagrams (e.g. the two diagrams with individuals arising from Euler diagram 1 in Table 1). Thus, no carry-over effect was anticipated from the clutter treatments.

Construct validity focuses on dependent variables (error rate, false negatives, and time) and independent variables (questions and treatments). Errors could arise if the diagrams were drawn in such a way that cognition was hindered (this could also increase time taken). To manage this effect, all diagrams were drawn to adhere to a set of layout guidelines, minimizing unwanted variation between them, and their labelling was carefully chosen.

To ensure the rigour of time measurements, consideration was paid to the duration elapsed interpreting a diagram as well as the units employed to measure time (60^{th}s of a second). The inclusion of a pause screen between each question ensured that the question was only displayed when the participant was ready to proceed. Further, participants used the same PC located in the same laboratory with no applications running in the background.

It was also considered a threat if the questions did not require sufficient cognitive effort to answer. To manage this, no two diagrams represented the same information: participants had to read each diagram before being able to answer the question. Moreover, this threat would arise if the diagrams were trivial: having only a few curves or zones, or individuals in only one set, was deemed insufficient to yield noticeable differences in performance, should they exist in general. So, the questions and the information conveyed ensured that the answer to the posed question involved three sets.

Regarding the clutter scores treatments, it was deemed a threat if there was no notable difference between high clutter and low clutter scores. The high clutter scores ranged from 13 to 17 and the low clutter scores ranged from 3 to 8, ensuring a distinct gap in scores between these two treatments.

Regarding the presence and absence treatments, it was deemed a threat if the high clutter scores arising from the use of presence and, respectively, absence were particularly dissimilar. Likewise for the low clutter scores. Due to the way in which the clutter score is calculated, it was not possible to ensure identical distributions of clutter scores across these two treatments, whilst also ensuring that participants saw semantically equivalent diagrams. Therefore, the score profiles were similar: high clutter presence scores were 15 and 17, paired with low clutter absence scores of 8 and 7, respectively. Similarly, the low clutter scores for presence, namely 3 and 5, were paired with high clutter scores of 14 and 13, respectively, when using absence. In each case, the absolute difference in the clutter scores is comparable, helping to control this threat to validity.

Lastly, we focus on external validity, by examining the limitations of the results and the extent to which they can be generalised. We observe the following: the low clutter scores ranged from 3 to 8 and the high clutter scores ranged from 13 to 17 and our diagrams included information about exactly one individual, in order to avoid visualizing redundant information. Further studies are needed to see if we get similar results for different clutter scores, diagrams containing multiple individuals, and for participants drawn from the general population. Thus, the results should be taken to be valid within these constraints.

8. General Discussion

Unsurprisingly, our results consistently indicated that, for all answer choices, the *low clutter presence diagrams* performed significantly better than *high* *clutter absence diagrams* (supporting H1, H3 and H5). To remind the reader, for answer choices 1 and 3, relating to H1 and H5 respectively, presence diagrams were well-matched unlike absence diagrams. Therefore, our results for H1 and H5 serve to support the validity of the existing theories that being well-matched [13] and low in clutter are important features of visualizations when performing tasks. Indeed, these results indicate that the theories are valid and are effective indicators of relative cognitive benefits. By contrast, for answer choice 2, presence and absence diagrams were both regarded to be well-matched, so their distinguishing feature was the level of clutter. Therefore, the acceptance of H3 indicated that being low in clutter—as suggested by research into visual search [10,19,42] and clutter [30]—is important for effective task performance.

Our results further indicated that for all three answer choices, the *high clutter presence diagrams* did not perform significantly differently to *low clutter absence diagrams.* With respect to answer choices 1 and 3, relating to H2 and H6, the results suggested that being well-matched is as important as being low in clutter, for diagrams of the type studied here. That is, if we have to choose between being well-matched or low in clutter then the choice made should not have a significant impact on task performance. Of course, we are mindful that our diagrams only represented one individual and it would be interesting to explore more fully whether the lack of one of these properties (e.g. not being well-matched) is compensated for by the existence of the other (e.g. being low in clutter), as the complexity of the diagrams increases. Moreover, exploring the same trade-off in other visual notations would be interesting—finding similar results elsewhere would provide a basis for positing general theories about how to choose between well-matchedness and clutter level in visualizations.

With respect to answer choice 2, both the presence and absence diagrams were regarded as being well-matched and a distinguishing feature was the level of clutter exhibited. This led us to hypothesis H4: *low clutter* absence diagrams would lead to significantly better task performance. However, no significant difference was revealed, suggesting that the *organization* [30] displayed by the high clutter presence diagrams—due to the connected nature of the trees representing the individuals—could have compensated for their high level of clutter. Our study was not designed to directly reveal whether the degree of organisation exhibited by the individuals impacted on task performance (we focused on clutter and individual type). It will, therefore, be interesting future work to empirically explore this further. To do so, one would need to define a measure that corresponds to the degree of organisation, likely to be relative to the connectivity displayed by the trees that represent individuals and the relative proximity of the regions containing nodes. Given such a measure, one could then run an empirical study using a similar method to that employed here, to determine any relationship with relative cognitive benefits. Lastly, whilst this result concerning H2 does not suggest the theories concerning clutter level and well-matchedness are invalid, they do indicate—unsurprisingly—that a multitude of features in a visualization contribute to its effectiveness as a representation of information.

The results of our study also suggested that further consideration may need to be given to the measure of clutter used. It could be that participants' performance is actually related to the *number of instances* of an individual's name. For instance, in Figure 5, the two diagrams each contain eight instances of Amy. The presence diagram has a high clutter score (15) whereas the absence diagram has a low clutter score (8); our results indicate that there would be no significant difference in task performance when using these two diagrams. By contrast, the presence diagram in Figure 4 contains just two instances of Paul (clutter score 3) whereas the absence diagram contains 14 instances of Paul (clutter score 14); here significant differences were seen. It may well be that the number of instances of an individual's name, rather than the clutter score proposed in earlier work, determines whether significant performance differences were observed. This could explain why, unexpectedly, no significant differences were found between high clutter presence and low clutter absence in *all three cases* (i.e. relating to H2, H4 and H6). In light of this observation, the results again do not invalidate the theory of well-matchedness as an important indicator of effectiveness, nor do they suggest that clutter is not important. Rather, they indicate that perhaps a different way of measuring clutter in these diagrams, more closely matched to the informational content than the syntactic structure of the diagrams, may be needed.

An important point is that no pairs of diagrams used in our study had a particularly smaller number of individual names for absence diagrams than for the 'partner' presence diagram. Thus, our results suggest a further study is needed to determine whether absence diagrams can allow people to perform significantly better than presence diagrams in this previously unconsidered case. If it is found that the number of individual names used in the diagram is the overriding feature that indicates relative task performance then this could imply that there is no relative difference in the degree of difficulty between interpreting disjunction (represented by connecting lines) in presence diagrams and negation (represented by over-lines) in absence diagrams. If so, this would support the inclusion of absence in diagrammatic logics generally. Moreover, it is known that people can experience more difficultly interpreting negated statements than their affirmative counterparts when using symbolic logics [22]. Therefore, it may be that diagrammatic logics provide a way to reduce the relative degree of difficultly associated with understanding negation.

9. Conclusion

It has been established that the way in which individuals are represented in diagrammatic logics can have a significant effect on human cognition. Through conducting an empirical study, we found that representing the absence of an individual in a way that yielded high diagram clutter significantly hindered task performance. By contrast, representing the presence of individuals, irrespective of the associated diagram clutter, or the absence of individuals in a low cluttered manner supports task performance. As a result, we are able to make empirically informed choices about how we represent individuals in diagrammatic logics. Thus, our research contributes to a major research problem faced by the diagrams community: to understand how to choose between semantically equivalent diagrams.

There is still much to be done, however, to address this problem. Our research, together with earlier research into Euler diagram effectiveness, guides us towards cognitively effective choices of monadic first-order diagrams. Further studies are needed to understand how to best represent other concepts, such as binary relations, needed to make semantically more complex statements. Research is also needed to understand the impact of syntactic choices on logical reasoning, beyond simply interpreting diagrams.

Acknowledgements. Thanks are due to Lopamudra Choudhury and Mihir Chakraborty for their helpful discussions on this research and to John Howse for his feedback on the study materials. This research was supported by EPSRC Grant EP/M011763/1 and, in part, conducted whilst Stapleton was at Jadavpur University as Visiting Professor.

References

[1] Alper, B., N. Henry Riche, G. Ramos, and M. Czerwinski, Design study of LineSets, a novel set visualization technique, *IEEE Transactions on Visualization and Computer Graphics* 17(12):2259–2267, 2011.

[2] Alqadah, M., G. Stapleton, J. Howse, and P. Chapman, Evaluating the impact of clutter in Euler diagrams, in *8th International Conference on the Theory and Application of Diagrams*, Springer, 2014, pp. 109–123.

[3] Blake, A., G. Stapleton, J. Howse, and P. Rodgers, The impact of topological and graphical properties on the perception of Euler diagrams, *Information Sciences* 330:455–482, 2015.

[4] Burton, J., M.K. Chakraborty, L. Chouduhry, and G. Stapleton, Minimizing clutter using absence in Venn-i^e, in *9th International Conference on the Theory and Application of Diagrams*, Springer, 2016, pp. 107–122.

[5] Chapman, P., G. Stapleton, P. Rodgers, L. Micallef, and A. Blake, Visualizing sets: An empirical comparison of diagram types, in *8th International Conference on the Theory and Application of Diagrams*, Springer, 2014, pp. 146–160.

[6] Choudhury, L., and M.K. Chakraborty, On extending Venn diagrams by augmenting names of individuals, in *3rd International Conference on the Theory and Application of Diagrams*, Springer, 2004, pp. 142–146.

[7] Chourdhury, L., and M.K. Chakraborty, On representing open universe, *Studies in Logic* 5(11):96–112, 2012.

[8] Chourdhury, L., and M.K. Chakraborty, Singular propositions, negation and the square of opposition, *Logica Universalis* 10(2):215–231, 2016.

[9] Dau, F., Constants and functions in Peirce's existential graphs, in *Conceptual Structures*, Springer, 2007, pp. 429–442.

[10] Eckstein, P.M., Visual search: A retrospective, *Journal of Vision* 11(5):14, 2011.

[11] Fish, A., and J. Masthoff, An experimental study into the default reading of constraint diagrams, in *IEEE Symposium Visual Languages and Human Centric Computing*, IEEE, 2005, pp. 287–289.

[12] Gil, J., J. Howse, and S. Kent, Formalising spider diagrams, in *IEEE Symposium on Visual Languages*, IEEE, 1999, pp. 130–137.

[13] Gurr, C., Effective diagrammatic communication: Syntactic, semantic and pragmatic issues, *Journal of Visual Languages and Computing* 10(4):317–342, 1999.

[14] Hammer, E., *Logic and visual information*, CSLI Publications, 1995.

[15] Healey, G.C., and J.T. Enns, Attention and visual memory in visualization and computer graphics, *IEEE transactions on visualization and computer graphics*, 18(7):1170–1188, 2012.

[16] Hou, T., P. Chapman, and A. Blake, Antipattern comprehension: An empirical evaluation, in *9th International Conference on Formal Ontology in Information Systems*, IOS Press, 2016, pp. 211–224.

[17] Howse, J., G. Stapleton, and J. Taylor, Spider diagrams, *LMS Journal of Computation and Mathematics* 8:145–194, 2005.

[18] Howse, J., G. Stapleton, K. Taylor, and P. Chapman, Visualizing ontologies: A case study, in *10th International Semantic Web Conference*, Springer, 2011, pp. 257–272.

[19] Huang, L., and H. Pashler, A Boolean map theory of visual attention, *Psychological review* 114(3):599–631, 2007.

[20] John, C., A. Fish, J. Howse, and J. Taylor, Exploring the notion of clutter in Euler diagrams, in *4th International Conference on the Theory and Application of Diagrams*, Springer, 2006, pp. 267–282.

[21] Kent, S., Constraint diagrams: Visualizing invariants in object oriented models, in *Proceedings of OOPSLA97*, ACM Press, 1997, pp. 327–341.

[22] Khemlani, S., I. Orenes, and P. Johnson-Laird, The negations of conjunctions, conditionals, and disjunctions, *Acta Psychologica* 151:1–7, 2014.

[23] Liang, K.Y., and S.L. Zeger, Longitudinal data analysis using generalized linear models, *Biometrika* 73:13–22, 1986.

[24] Meulemans, W., N. Henry Riche, B. Speckmann, B. Alper, and T. Dwyer, Kelpfusion: A hybrid set visualization technique, *IEEE Transactions on Visualization and Computer Graphics* 19(11):1846–1858, 2013.

[25] Mineshima, K., M. Okada, and R. Takemura, A diagrammatic inference system with Euler circles, *Journal of Logic, Language and Information* 21(3):365–391, 2012.

[26] Moktefi, A., Is Euler's circle a symbol or an icon?, *Sign Systems Studies* 43(4):597–615, 2015.

[27] Purchase, H.C., *Experimental Human Computer Interaction: A practical guide with visual examples*, Cambridge University Press, 2012.

[28] Riche, N., and T. Dwyer, Untangling Euler diagrams, *IEEE Transactions on Visualization and Computer Graphics* 16(6):1090–1099, 2010.

[29] Rodgers, P., L. Zhang, and H. Purchase, Wellformedness properties in Euler diagrams: Which should be used?, *IEEE Transactions on Visualization and Computer Graphics* 18(7):1089–1100, 2012.

[30] Rosenholtz, R., Y. Li, and L. Nakano, Measuring visual clutter, *Journal of Vision* 7(2):17, 2007.

[31] Sato, Y., S. Masuda, Y. Someya, T. Tsujii, and S. Watanabe, An fMRI analysis of the efficacy of Euler diagrams in logical reasoning, in *IEEE Symposium on Visual Languages and Human-Centric Computing*, 2015, 2015, pp. 143–151.

[32] Sato, Y., and M. Mineshima, How diagrams can support syllogistic reasoning: An experimental study, *Journal of Logic, Language and Information* 24(4):409–455, 2015.

[33] Shin, S.-J., *The logical status of diagrams*, Cambridge University Press, 1994.

[34] Shin, S.-J., *The iconic logic of Peirce's graphs*, Bradford Book, 2002.

[35] Stapleton, G., A. Blake, J. Burton, and A. Touloumis, Supplementary material, http://readableproofs.org/absence-study-2016. Accessed September 2016.

[36] Stapleton, G., J. Howse, S. Thompson, J. Taylor, and P. Chapman, On the completeness of spider diagrams augmented with constants, in S.-J. Shin, and A. Moktefi, (eds.), *Visual Reasoning with Diagrams*, Birkhauser, 2013, pp. 107–122.

[37] Stapleton, G., J. Taylor, J. Howse, and S. Thompson, The expressiveness of spider diagrams augmented with constants, *Journal of Visual Languages and Computing* 20:30–49, 2009.

[38] Swoboda, N., Implementing Euler/Venn reasoning systems, in M. Anderson, B. Meyer, and P. Olivier, (eds.), *Diagrammatic representation and reasoning*, Springer-Verlag, 2001, pp. 371–386.

[39] Swoboda, N., and G. Allwein, Using DAG transformations to verify Euler/Venn homogeneous and Euler/Venn FOL heterogeneous rules of inference, *Journal on Software and System Modeling* 3(2):136–149, 2004.

[40] Szymanik, J., and R. Verbrugge, Logic and cognition: Special issue of best papers of the ESSLLI 2012 workshop, *Journal of Logic, Language and Information* 22:357–362, 2013.

[41] Takemura, R., Counter-example construction with Euler diagrams, *Studia Logica* 103:669–696, 2015.

[42] Wolfe, J.M., Guided Search 2.0 A revised model of visual search, *Psychonomic Bulletin & Review* 1(2):202–238, 1994.

G. Stapleton, A. Blake, J. Burton, A. Touloumis
University of Brighton
Brighton
UK
g.e.stapleton@brighton.ac.uk

A. Blake
a.l.blake@brighton.ac.uk

J. Burton
j.burton@brighton.ac.uk

A. Touloumis
A.Touloumis@brighton.ac.uk