Minimizing Clutter Using Absence in Venn-i[e]

Jim Burton[1], Mihir Chakraborty[2], Lopamudra Choudhury[2], and Gem Stapleton[1(✉)]

[1] Visual Modelling Group, University of Brighton, Brighton, UK
{j.burton,g.e.stapleton}@brighton.ac.uk
[2] Jadavpur University, Kolkata, India
mihirc4@gmail.com, choudhuryl@yahoo.com

Abstract. Over the last two decades substantial advances have been made in our understanding of diagrammatic logics. Many of these logics have the expressiveness of monadic first-order logic, sometimes with equality, and are equipped with sound and complete inference rules. A particular challenge is the representation of *negated* statements. This paper addresses the problem of how to represent negated statements involving constants, thus asserting the absence of specific individuals, in the context of Euler-diagram-based logics. Our first contribution is to explore the potential benefits of explicitly representing absence using constants, in terms of clutter reduction, and to highlight ontological issues that arise. We go on to define a measure of clutter arising from constants. By defining a set of semantics-preserving inference rules, we are able to algorithmically minimize diagram clutter, in part made possible by the inclusion of absence. Consequently, information about individuals can be represented in a minimally cluttered way.

1 Introduction

Negation, closely related to the notion of absence, plays a crucial role in all logics. Indeed, "The capacity to negate is the capacity to refuse, to contradict, to lie, to speak ironically, to distinguish truth from falsity – in short, the capacity to be human" [8]. It has long been recognized that diagrams are sometimes unable to explicitly represent negated statements. Indeed, many of the logics based on Euler diagrams do not permit statements such as $a \notin P$ to be made explicitly. Instead, one has to assert that $a \in P'$, where P' is the complement of P. There is, however, one exception to this: Choudhury and Chakraborty developed a classical logic called Venn-i that allows $a \notin P$ to be *directly* expressed [5].

Venn-i extends Shin's Venn-I system, which includes Peirce's ⊗-sequences to assert non-emptiness of sets [13], alongside i-sequences and $\bar{i}$-sequences to represent individuals and their absence. Since Choudhury and Chakraborty adopt a classical interpretation, the absence of an individual from one set implies its presence in the complement. In Fig. 1, D_1 uses an i-sequence to assert $a \in P \backslash Q$, using an a, and D_2 negates this statement, expressing $a \notin P \backslash Q$, using an $\bar{i}$-sequence. Moreover, D_2 is semantically equivalent to D_3, which expresses

M. Jamnik et al. (Eds.): Diagrams 2016, LNAI 9781, pp. 107–122, 2016.
DOI: 10.1007/978-3-319-42333-3_9

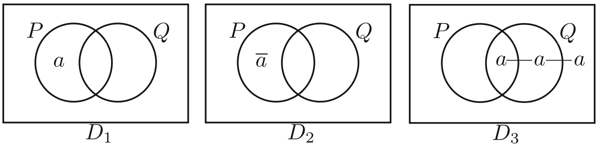

Fig. 1. Asserting presence and absence.

$a \in (P \cap Q) \cup (Q \backslash P) \cup ((U \backslash P) \cap (U \backslash Q))$ using an i-sequence, namely $a - a - a$. An inspiration for Choudhury's and Chakraborty's work came from the notion of *abhāva* (absence). Abhāva, an important feature of ancient Indian knowledge systems, allocates a first class status to the absence of individuals. A philosophical account of absence can be found in [4].

Speaking from the point of view of cognitive science, absence would indicate that though we do not directly perceive the object, we do perceive its absence; there is a mental imagery of the absent object. Thus, when considering a particular individual (of which we have a mental image) we check whether it is in a particular locus and directly perceive its absence. This is reflected by the treatment of Venn-i as a classical logic, where the law of excluded middle holds, as opposed to a sort of constructivist logic where the absence of an individual from one set need not imply its presence in the complement [3].

As we will demonstrate, explicitly representing the absence of individuals allows information to be presented in a less cluttered way. Clutter in Euler diagrams, which are closely related to Venn diagrams, was studied by John et al. [11]: they devised a theoretical measure of clutter. Alqadah et al. established that increased levels of clutter in Euler diagrams negatively impacts user task performance [1]. Hence, there is clearly a need to theoretically understand clutter in diagrams generally and its impact on end-user task performance.

This paper takes the first step towards understanding clutter arising from the sequences in an extended version of Venn-i, which we call Venn-i^e, by:

- Discussing the interplay between absence and presence, as well as highlighting their asymmetry (Sect. 2),
- Formalizing the syntax and semantics of Venn-i^e, which use Euler diagrams as a basis[1] (Sect. 3),
- Defining a measure of clutter arising from $\otimes$-sequences, i-sequences and $\bar{i}$-sequences (Sect. 4); we note here that i-sequences can comprise many nodes whereas $\bar{i}$-sequences always have a single node,
- Identifying necessary and sufficient conditions for Venn-i^e diagrams to be unsatisfiable (Sect. 5),

[1] Using Euler diagrams as a basis renders Swoboda's and Allwein's Euler/Venn logic, which does not include $\bar{i}$-sequences, a proper fragment of Venn-i^e. Indeed, many techniques that have been devised for visualizing sets extend Euler diagrams (or variations of them) by the inclusion of individuals, such as [6,12,14]. Whilst not viewed as logics, our work is relevant to these systems since absence provides an alternative way of asserting the set to which an individual belongs.

– Demonstrating how to minimize clutter in satisfiable diagrams by defining inference rules for altering sequences (Sect. 6), and
– Discussing the role of absence in clutter reduction and its potential implications on task performance (Sect. 7).

We conclude and discuss future work in Sect. 8.

2 Representing Absence Diagrammatically

Semantically equivalent statements can be made about the sets in which an individual lies using either *positive* or *negative* statements, such as $a \in P \cup Q$ versus $a \notin P' \cap Q'$ respectively. Whilst various diagrammatic logics include syntax to explicitly make positive statements like $a \in P \cup Q$, including [18, 19], they have overlooked the possibility of making negative statements like $a \notin P' \cap Q'$. One benefit of allowing diagrams to make negative statements is that less cluttered diagrams can be formed: using $\overline{a}$ signs in diagrams can be more succinct, relative to diagrams using a signs; see Fig. 2. As previously noted, clutter can have a significant negative impact on diagram comprehension.

In Fig. 3, the three diagrams are semantically equivalent. We can reduce the clutter in D_1 by substituting $\overline{a}$ for the a-sequence, with the result shown in D_2. As well as swapping syntax that makes positive (resp. negative) statements for syntax that makes negative (resp. positive) statements, clutter can also be reduced by removing redundant syntax. The diagram D_3 has more syntax than D_1, such as two additional $\otimes$-sequences, and is more cluttered as a result.

There are fundamental ontological differences between pieces of syntax representing presence and absence. This is because, although syntactically similar, the semantic status of a and $\overline{a}$ signs is different. Firstly, there are differences relating to their locations within a diagram. If there are distinct i-sequences with the same label placed in disjoint regions then the diagram is inconsistent: disjoint regions represent disjoint sets and a given individual cannot be in two disjoint sets. By contrast, $\overline{a}$ can be placed in several disjoint regions without giving rise to inconsistency per se: it is entirely possible for an individual to be absent from two disjoint sets, for instance.

Secondly, we observe that the presence of a sequence, either of the form a or $\overline{a}$, in some region, r, carries existential import. However, this existential import behaves differently: we see that a drawn inside r implies the set, s, that r represents is not empty, whereas $\overline{a}$ drawn in r implies the complement of s is not

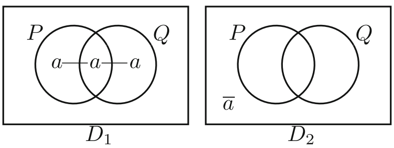

Fig. 2. Making positive (left) and negative (right) statements.

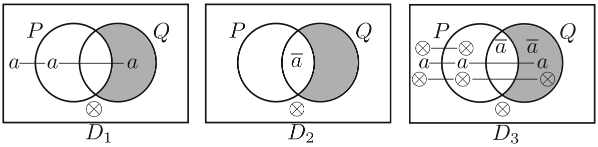

Fig. 3. Diagrams with different levels of clutter.

empty. Thus, the role of absence in terms of existential import is asymmetrical with presence. This may affect the way diagrams are understood by users.

Thirdly, the interaction of absence with subsumption may contradict intuition. In Fig. 4, D_1 tells us that $Q \subseteq P$ and $a \notin Q$. However, this does not imply $a \notin P$, so D_1 does not imply D_2; by contrast, $Q \subseteq P$ and $a \in Q$ implies $a \in P$. This behaviour runs counter to the iconicity [17] of Euler diagrams, which are known to support inference through mechanisms such as free rides [15]. Iconicity is exploited in Euler diagrams through the way that containment indicates subsumption: elements that belong to a set represented by a contained circle belong, "naturally", to the set represented by the containing circle. On the other hand, the absence of an individual from a set represented by a contained circle does not imply absence from the set represented by the containing circle. Thus, with regard to subsumption, $\overline{a}$ does not behave transitively, unlike a.

To summarize, explicitly representing the absence of individuals allows clutter to be reduced in diagrams. Moreover, we must be mindful of various ontological differences between a and $\overline{a}$ when reasoning.

3 Syntax and Semantics of Venn-i^e

Venn-i^e extends Venn-i introduced in [5], relaxing the restriction to Venn diagrams by allowing Euler diagrams to be used. In turn, Venn-i extends Shin's Venn-I system [16]. As is typical, the abstract syntax is given alongside an informal description of the concrete syntax.

Consider the Venn-i^e diagram in Fig. 5. There are two closed curves, labelled P and Q. We conflate the closed curves with their labels and simply say 'the curve P', or just 'P'. The curves give rise to three *zones*: a zone is a region inside some (possibly no) curves and outside the remaining curves. In Fig. 5, the only *shaded* zone is inside Q but outside P. The diagram also contains four graphs:

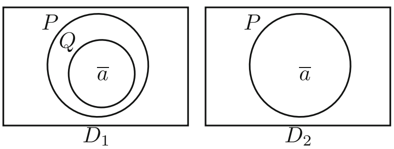

Fig. 4. The interplay between absence and subsumption.

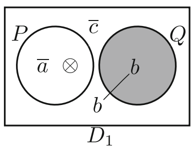

Fig. 5. Syntax: Venn-i^e.

1. One $\otimes$-sequence which comprises a single node,
2. One i-sequence (i for individual), namely b, comprising two nodes joined by one edge, and
3. Two $\bar{i}$-sequences, namely $\bar{a}$ and $\bar{c}$, both of which comprise a single node.

Typically, the abstract syntax for an Euler diagram, D, comprises a set of labels, a set of zones, and a set of shaded zones, written $D = (L, Z, ShZ)$. Zones are ordered pairs of finite, disjoint sets of labels, (in, out), where in (resp. out) denotes the (labels of) the curves that the zone is inside (resp. outside). The zone outside all of the curves, namely $(\emptyset, L)$, must be in D and any zone in D satisfies $in \cup out = L$. The set ShZ of shaded zones only contains zones in Z.

In Fig. 5, the underlying Euler diagram is (L, Z, ShZ), where $L = \{P, Q\}$, $Z = \{(\emptyset, \{P, Q\}), (\{P\}, \{Q\}), (\{Q\}, \{P\})\}$ and $ShZ = \{(\{Q\}, \{P\})\}$. The zone $(\emptyset, \{P, Q\})$ is that which is outside all of the curves, hence the first part of the ordered pair being $\emptyset$ and the second part containing both P and Q. As we shall see, this zone denotes the set $P' \cap Q'$.

It is helpful for us to have a set of labels from which all labels used in any diagram are drawn; we call this set $\mathcal{L}$. When making general statements, we take $\mathcal{L} = \{\lambda_1, \lambda_2, ...\}$ whereas in examples we use P, Q, R, and so forth. Given $\mathcal{L}$, the set of all zones is denoted $\mathcal{Z}$. We also have a set of constant symbols, denoted $\mathcal{C}$, which gives rise to i-sequences and $\bar{i}$-sequences. We take $\mathcal{C} = \{\iota_1, \iota_2, ...\}$; in examples, we use a, b, c, and so forth.

The *regions* (i.e. non-empty sets of zones) in a diagram need to be associated with the sequences drawn in them. In general, $\otimes$-sequences and i-sequences can have nodes placed in many zones, whereas $\bar{i}$-sequences always have a single node. This reflects the dual role of i-sequences and $\bar{i}$-sequences: an a-sequence in the region $\{z_1, ..., z_n\}$ asserts that $a \in z_1 \vee ... \vee a \in z_n$ which is equivalent to $a \notin z_{n+1} \wedge ... \wedge a \notin z_{n+m}$, where $z_{n+1}, ..., z_{n+m}$ are the zones not in r. This equivalent statement can be made by a set of $\bar{a}$-sequences, one in each zone not in r. To identify the sequences in each region, we use three binary relations $\rho_\otimes$, ρ_i and $\rho_{\bar{i}}$. In Fig. 5, $\rho_\otimes = \{(\{(\{P\}, \{Q\})\}, \otimes_1)\}$, $\rho_i = \{(\{(\{Q\}, \{P\}), (\emptyset, \{P, Q\})\}, b)\}$, and $\rho_{\bar{i}} = \{((\{P\}, \{Q\}), a), ((\emptyset, \{P, Q\}), c)\}$.

Definition 1. *A **Venn-i^e** diagram, D, is a tuple, $D = (L, Z, ShZ, \rho_\otimes, \rho_i, \rho_{\bar{i}})$ such that:*

1. *L is a finite set of labels chosen from $\mathcal{L}$.*
2. *Z is a set of zones where $(\emptyset, L) \in Z$ and for all (in, out) in Z, $in \cup out = L$.*
3. *ShZ is a subset of Z whose elements are called **shaded** zones.*
4. *$\rho_\otimes \subseteq (\mathbb{P}Z \backslash \{\emptyset\}) \times \{\otimes\}$ is a finite binary relation that associates non-empty regions with $\otimes$ symbols. The elements of $\rho_\otimes$ are called **$\otimes$-sequences**.*
5. *$\rho_i \subseteq (\mathbb{P}Z \backslash \{\emptyset\}) \times \mathcal{C}$ is a finite binary relation that associates non-empty regions with constant symbols. The elements of ρ_i are called **i-sequences**.*
6. *$\rho_{\bar{i}} \subseteq Z \times \mathcal{C}$ is a finite binary relation that associates zones with constant symbols. The elements of $\rho_{\bar{i}}$ are called **$\bar{i}$-sequences**.*

*The **missing zones** of D are elements of $MZ = \{(in, out) \in \mathcal{Z} : in \cup out = L\} \backslash Z$. Furthermore, given a constant, ι, the set of ι-sequences in D is denoted*

$I(\iota)$ where $I(\iota) = \{(r, \iota) : (r, \iota) \in \rho_i\}$. Similarly, the set of $\bar{\iota}$-sequences in D is denoted $I(\bar{\iota})$ where $I(\bar{\iota}) = \{(z, \iota) : (z, \iota) \in \rho_{\bar{i}}\}$.

The underlying Euler diagrams have the typical semantics: the closed curves represent sets and their spatial relationships correspond to set-theoretic relationships. Shading asserts emptiness, as seen in Shin's systems [16]. Sequences give information about the location of elements in sets. First, $\otimes$-sequences, introduced by Peirce [13], assert the non-emptiness of sets. Second, i-sequences assert that the denoted individuals are in the sets represented by the regions in which they are placed. Lastly, each $\bar{i}$-sequence asserts the absence of the denoted individual from the set represented by the zone in which it is placed. In Fig. 5, the b-sequence asserts that $b \in Q \cap P'$ *or* $b \in P' \cap Q'$, since b is in the two zone region $\{(\{Q\}, \{P\}), (\emptyset, \{P, Q\})\}$. Likewise, the $\bar{c}$-sequence is in the zone $(\emptyset, \{P, Q\})$ which means that $c \notin P' \cap Q'$. To formalize the semantics, we adopt a standard model-theoretic approach.

Definition 2. *An* ***interpretation****, $\mathcal{I}$, is a triple, $\mathcal{I} = (U, \psi, \Psi)$, such that*

1. *U is a non-empty set, called the* ***universal set****,*
2. *$\psi\colon \mathcal{C} \to U$ maps constants to elements in U, and*
3. *$\Psi\colon \mathcal{L} \to \mathbb{P}U$ maps curve labels to subsets of U.*

The function Ψ is extended to interpret zones as follows: for each zone, (in, out),

$$\Psi(z) = \bigcap_{l \in in} \Psi(l) \cap \bigcap_{l \in out} (U \backslash \Psi(l)).$$

Definition 3. *Let $D = (L, Z, ShZ, \rho_\otimes, \rho_i, \rho_{\bar{i}})$ be a Venn-i^e diagram and let $\mathcal{I} = (U, \psi, \Psi)$ be an interpretation. Then $\mathcal{I}$ is a* ***model*** *for D provided the following conditions all hold.*

1. ***Missing Zones Condition:*** *for each $z \in MZ$, $\Psi(z) = \emptyset$.*
2. ***Shaded Zones Condition:*** *for each $z \in ShZ$, $\Psi(z) = \emptyset$.*
3. ***$\otimes$-Sequence Condition:*** *for each $(r, \otimes) \in \rho_\otimes$, $\Psi(z) \neq \emptyset$ for some $z \in r$.*
4. ***i-Sequence Condition:*** *for each $(r, \iota) \in \rho_i$, $\psi(\iota) \in \Psi(z)$ for some $z \in r$.*
5. **$\bar{i}$-Sequence Condition:** *for each $(z, \iota) \in \rho_{\bar{i}}$, $\psi(\iota) \notin \Psi(z)$.*

If $\mathcal{I}$ models D then $\mathcal{I}$ ***satisfies*** *D. Diagrams with no models are* ***unsatisfiable****.*

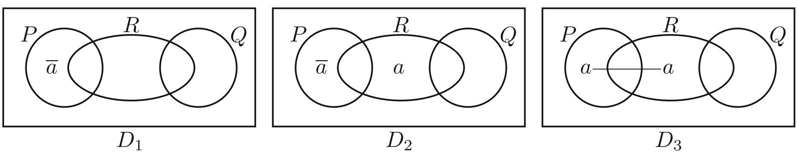

Fig. 6. Measuring clutter in Venn-i^e diagrams.

4 Measuring Clutter

We require a measure of clutter arising from the sequences. Figure 6 shows three simple examples, all with the same underlying Euler diagram. The lefthand diagram with just one node, namely $\overline{a}$, is less cluttered than the middle diagram. The righthand diagram is the most cluttered, since this has two nodes (both named a) *and* a connecting edge. Thus, to measure the clutter arising from the sequences, we count the number of nodes and the number of edges.

Definition 4. *Let* $D = (L, Z, ShZ, \rho_{\otimes}, \rho_i, \rho_{\overline{i}})$ *be a Venn-i^e diagram. The* ***sequence clutter score*** *for* D*, denoted* $\mathcal{SCS}(D)$*, is*

$$\mathcal{SCS}(D) = \left(\sum_{(r,\otimes)\in\rho_{\otimes}} (2|r| - 1) \right) + \left(\sum_{(r,\iota)\in\rho_i} (2|r| - 1) \right) + |\rho_{\overline{i}}|$$

The three diagrams in Fig. 6 have sequence clutter scores 1, 2, and 3 respectively. From this point forward, we simply say *clutter score*.

Definition 5. *Let* $D_1 = (L, Z, ShZ, \rho_{\otimes}, \rho_i, \rho_{\overline{i}})$ *be a Venn-i^e diagram. Then* D *is* ***minimally cluttered*** *if there does not exist a semantically equivalent diagram,* $D' = (L, Z, ShZ, \rho'_{\otimes}, \rho'_i, \rho'_{\overline{i}})$*, such that* $\mathcal{SCS}(D') < \mathcal{SCS}(D)$*.*

5 Minimizing Clutter in Inconsistent Diagrams

Figure 7 shows a minimally cluttered inconsistent diagram, namely D_1: it has a clutter score of 0; thus, any inconsistent diagram is semantically equivalent to D_1. To allow us to focus on consistent diagrams, when algorithmically reducing clutter, we need to identify syntactic conditions which capture inconsistency. There are various ways in which Venn-i^e diagrams can be inconsistent:

1. All interpretations have a non-empty universal set, so a diagram is inconsistent if it is entirely shaded. See D_1 in Fig. 7.
2. Shaded regions containing entire $\otimes$-sequences or i-sequences are inconsistent since the shading asserts set emptiness whereas the sequence implies set non-emptiness. See D_2 in Fig. 7, where each sequence gives rise to inconsistency.
3. There are i-sequences placed in regions that do not share a common non-shaded zone, z, where z does not contain an $\overline{a}$-sequence. Intuitively, for each ι, the individual represented must lie in the set denoted by a non-shaded zone that is shared by all ι-sequences in $I(\iota)$. If all such zones include $\overline{\iota}$ then the diagram also asserts that the individual is absent from the sets represented by those zones. See D_3 in Fig. 7.

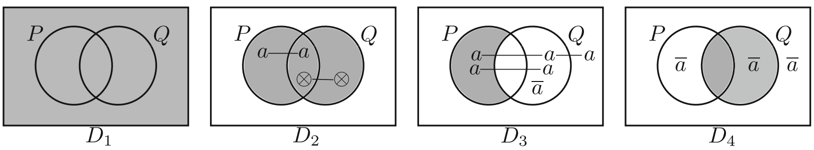

Fig. 7. Inconsistent Venn-i^e diagrams.

4. The set of $\bar{i}$-sequences for constant symbol ι, namely $I(\bar{\iota})$, cannot include all non-shaded zones. If all non-shaded zones were included then the law of excluded middle tells us that the represented individual must be an element of the empty set. See D_4 in Fig. 7.

Definition 6 (Inconsistency). *Let* $D = (L, Z, ShZ, \rho_{\otimes}, \rho_i, \rho_{\bar{i}})$ *be a Venn-*i^e *diagram. Whenever any one of the following conditions holds D is* ***inconsistent****.*

1. *All zones are shaded:* $Z = ShZ$*.*
2. *There is an* $\otimes$*-sequence, say* $(r, \otimes)$*, in* D *such that* $r \subseteq ShZ$*.*
3. *There is an* i*-sequence, say* (r, ι)*, in* D *such that for all zones,* z*, in* $\bigcap_{r' \in I(\iota)} r'$*, either* z *is shaded or* $(z, \iota) \in I(\bar{\iota})$*.*
4. *There is an* $\bar{i}$*-sequence, say* (z, ι)*, in* D *such that* $Z \backslash \{z' : (z', \iota) \in I(\bar{\iota})\} \subseteq ShZ$*.*

If D is not inconsistent then D is ***consistent****.*

Theorem 1 (Inconsistent). *D is inconsistent iff D is unsatisfiable.*

Using Theorem 1 we can therefore identify whether any given diagram is inconsistent. Given such a diagram $D = (L, Z, Z, \rho_{\otimes}, \rho_i, \rho_{\bar{i}})$ we can see that a minimally cluttered, semantically equivalent diagram is $D_{min} = (L, Z, Z, \emptyset, \emptyset, \emptyset)$.

6 Minimizing Clutter in Consistent Diagrams

The goal of this section is to produce minimally cluttered diagrams using inference rules that alter their sequences. To this end, we first define some useful transformations on diagrams.

Transformation 1 (Sequence Removal). *Let* $D = (L, Z, ShZ, \rho_{\otimes}, \rho_i, \rho_{\bar{i}})$ *be a Venn-*i^e *diagram. Let* $(r, \bullet)$ *be a sequence in* D*. We define three removal operations on* D*:*

1. *If* $(r, \bullet) \in \rho_{\otimes}$ *then* $D - (r, \bullet) = (L, Z, ShZ, \rho_{\otimes} \backslash \{(r, \bullet)\}, \rho_i, \rho_{\bar{i}})$*.*
2. *If* $(r, \bullet) \in \rho_i$ *then* $D - (r, \bullet) = (L, Z, ShZ, \rho_{\otimes}, \rho_i \backslash \{(r, \bullet)\}, \rho_{\bar{i}})$*.*
3. *If* $(r, \bullet) \in \rho_{\bar{i}}$ *then* $D - \overline{(r, \bullet)} = (L, Z, ShZ, \rho_{\otimes}, \rho_i, \rho_{\bar{i}} \backslash \{(r, \bullet)\})$*.*

Transformation 2 (Sequence Addition). *Let* $D = (L, Z, ShZ, \rho_{\otimes}, \rho_i, \rho_{\bar{i}})$ *be a Venn-*i^e *diagram. Let* $(r, \bullet)$ *be a sequence such that* $r \subseteq Z$ *or* $r \in Z$*. We define three addition operations on* D*:*

1. *If* $r \subseteq Z$ *and* $\bullet = \otimes$ *then* $D + (r, \bullet) = (L, Z, ShZ, \rho_{\otimes} \cup \{(r, \bullet)\}, \rho_i, \rho_{\bar{i}})$*.*
2. *If* $r \subseteq Z$ *and* $\bullet \in \mathcal{C}$ *then* $D + (r, \bullet) = (L, Z, ShZ, \rho_{\otimes}, \rho_i \cup \{(r, \bullet)\}, \rho_{\bar{i}})$*.*
3. *If* $r \in Z$ *and* $\bullet \in \mathcal{C}$ *then* $D + \overline{(r, \bullet)} = (L, Z, ShZ, \rho_{\otimes}, \rho_i, \rho_{\bar{i}} \cup \{(z, \bullet)\})$*.*

Before we present our inference rules, we work through an example showing how to minimize clutter. Consider Fig. 8. Here, the diagram D is consistent, but not minimally cluttered. To reduce clutter, we make various observations and adopt the following process:

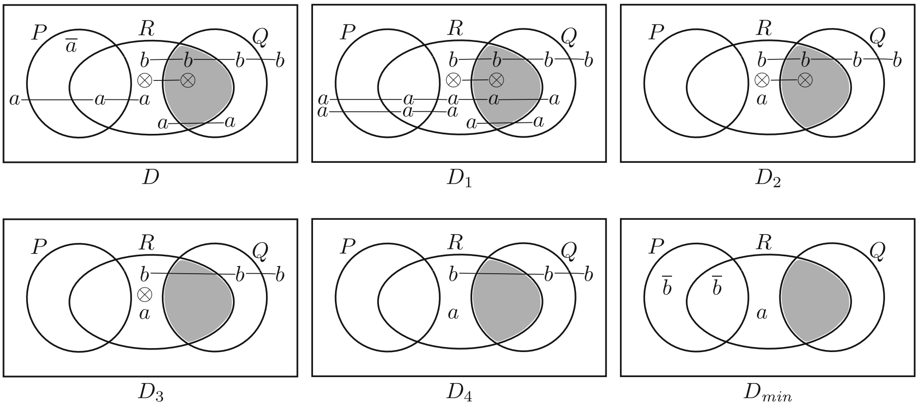

Fig. 8. Clutter reduction in consistent diagrams.

1. First we observe that whenever we express information using $\bar{i}$-sequences, we can instead use an i-sequence. Thus, in D we can swap $\overline{a}$ for an a-sequence, as shown in D_1. In general, this swap may result in the clutter score increasing, but it allows us to more easily identify, syntactically, the region in which a must represent an element.
2. Next, we observe that in D_1 (and, in any diagram), we only need one occurrence of each constant symbol to specify in which set it lies. So, we can reduce the three a-sequences in D_1 to a single a-sequence shown in D_2. This single a-sequence is placed in the zone common to all of the a-sequences in D_1, thus allowing us to see which region contains the individual a. In this step, the clutter score of D_2 is lower than that of D_1.
3. Reductions can also be made to sequences that are placed in regions which contain shaded zones, since shaded zones represent empty sets. The diagram D_2 contains two such sequences, (b, r_b) and $(\otimes, r_\otimes)$, and can be replaced by D_3.
4. Some sequences can be redundant from diagrams. In D_3, the $\otimes$-sequence is redundant since it tells us that $Q \backslash (P \cup R) \neq \emptyset$ which can be deduced from the a-sequence. So D_3 can be replaced by D_4.
5. Lastly, we examine each i-sequence in turn. If its contribution to the clutter score can be reduced by swapping it for $\bar{i}$-sequences then this swap is performed. Here, the b-sequence is swapped for two $\overline{b}$-sequences, resulting in D_{min}. This last step exploits the use of absence to reduce diagram clutter.

As we have just seen, it is possible to swap i-sequences for $\bar{i}$-sequences, and vice versa, reflecting their dual roles. For example, in Fig. 9, the a-sequence in D_1 tells us $a \in P \cap Q' \cap R'$ or $a \in P \cap Q \cap R'$. Given the shading and the spatial relationships between the curves, asserting $a \notin P' \cap Q' \cap R'$ is equivalent. This alternative representation is seen in D_2. We can *swap* the i-sequence $(\{(\{P\}, \{Q, R\}), (\{P, Q\}, \{R\})\}, a)$ for the $\bar{i}$-sequence $((\emptyset, \{P, Q, R\}), a)$.

Inference Rule 1 (Swap i-Sequence). *Let $D = (L, Z, ShZ, \rho_{\otimes}, \rho_i, \rho_{\bar{i}})$ be a Venn-i^e diagram. Let (r, ι) be an i-sequence in D. Then (r, ι) may be swapped for the set $\{(z, \iota) : z \in Z \backslash (ShZ \cup r)\} = \{\overline{(z_1, \iota)}, ..., \overline{(z_n, \iota)}\}$ of $\bar{i}$-sequences. That is, D may be replaced by $D - (r, \iota) + \overline{(z_1, \iota)} + ... + \overline{(z_n, \iota)}$ and vice versa.*

Inference Rule 2 (Swap $\bar{i}$-Sequences). *Let $D = (L, Z, ShZ, \rho_{\otimes}, \rho_i, \rho_{\bar{i}})$ be a Venn-i^e diagram. Let ι be a constant symbol such that $I(\bar{\iota}) \neq \emptyset$ and $Z \backslash (ShZ \cup \{z_1, ..., z_n\}) \neq \emptyset$, where $I(\bar{\iota}) = \{(z_1, \iota), ..., (z_n, \iota)\}$. Then $I(\bar{\iota})$ may be swapped for the i-sequence $(Z \backslash (ShZ \cup \{z_1, ..., z_n\}), \iota)$. That is, D may be replaced by*

$$D - \overline{(z_1, \iota)} - ... - \overline{(z_n, \iota)} + (Z \backslash (ShZ \cup \{z_1, ..., z_n\}), \iota)$$

and vice versa.

There are also occasions when we can remove parts of sequences: when the region in which a sequence is placed includes a shaded zone, the part in the shaded zone can be deleted, thus reducing the sequence. Moreover, we have also seen that sets of i-sequences can be reduced.

Inference Rule 3 (Reduce Sequence). *Let $D = (L, Z, ShZ, \rho_{\otimes}, \rho_i, \rho_{\bar{i}})$ be a Venn-i^e diagram. Let $(r, \bullet)$ be a sequence in D such that r contains at least two zones, one of which, z say, is shaded. Then D may be replaced by $D - (r, \bullet) + (r \backslash \{z\}, \bullet)$ and vice versa. Such a sequence is said to be* ***reducible*** *in D.*

Inference Rule 4 (Reduce a Set of Sequences). *Let $D = (L, Z, ShZ, \rho_{\otimes}, \rho_i, \rho_{\bar{i}})$ be a Venn-i^e diagram. Let ι be a constant symbol such that $I(\iota) \neq \emptyset$ and $r \neq \emptyset$ where*

$$r = \Big(\bigcap_{(r_i, \iota) \in I(\iota)} r_i \Big) \backslash \Big(ShZ \cup \bigcup_{(z, \iota) \in I(\bar{\iota})} \{z\} \Big).$$

Then D may be replaced by

$$D - (r_1, \iota) - ... - (r_n, \iota) + (r, \iota)$$

and vice versa, where $I(\iota) = \{(r_1, \iota), ..., (r_n, \iota)\}$. The set $I(\iota)$ of sequences is said to be ***reducible*** *in D.*

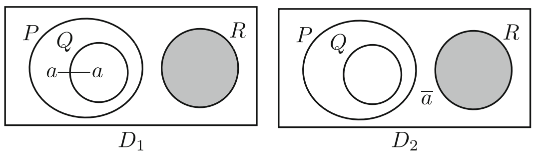

Fig. 9. Swapping sequences.

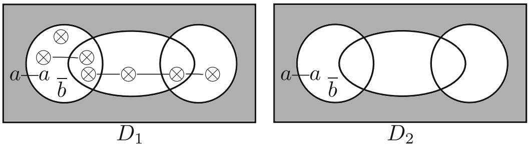

Fig. 10. Redundant ⊗-sequences.

There are various ways in which an ⊗-sequence can be redundant in a diagram, in the sense that its removal does not alter the semantics:

1. An ⊗-sequence, $(r, \otimes)$, that includes all of the non-shaded zones in r is redundant, since this amounts to asserting that $U \neq \emptyset$ which is necessarily true in all interpretations.
2. In D_1, Fig. 10, the single-node ⊗-sequence asserts $P \cap Q' \cap R' \neq \emptyset$. From this we can deduce $P \neq \emptyset$, asserted by the two-node ⊗-sequence which is, thus, redundant.
3. In D_1, the a-sequence tells us that $a \in P \cap Q' \cap R'$ or $a \in P' \cap Q' \cap R'$. The shading asserts $P' \cap Q' \cap R' = \emptyset$, so $a \in P \cap Q' \cap R'$. This implies that $P \cap Q' \cap R' \neq \emptyset$, so the single-node ⊗-sequence is also redundant. The *presence* of the individual a has permitted a reduction in diagram clutter.
4. Lastly, in D_1 the location of $\bar{b}$ tells us that $b \notin P \cap Q' \cap R'$, from which – together with the shading – it follows that $b \in Q \cup R$. Therefore, the four-node ⊗-sequence asserting $Q \cup R \neq \emptyset$ is redundant. The *absence* of the individual b has permitted a reduction in diagram clutter.

Removing the ⊗-sequences from D_1 in Fig. 10 to give D_2 reduces the clutter score from 15 to 4.

Inference Rule 5 (Remove ⊗-Sequence). *Let $D_1 = (L, Z, ShZ, \rho_\otimes, \rho_i, \rho_{\bar{i}})$ be a Venn-i^e diagram and let $(r, \otimes)$ be an ⊗-sequence in D such that either:*

1. *The region r includes all non-shaded zones: $Z \backslash ShZ \subseteq r$,*
2. *There is a distinct ⊗-sequence, say $(r', \otimes)$, in D where $r' \backslash ShZ \subseteq r$,*
3. *There is an i-sequence, say (r', ι), in D such that $r' \backslash ShZ \subseteq r$, or*
4. *Given $I(\bar{\iota}) = \{(z_1, \iota), ..., (z_n, \iota)\}$, it is the case that $(Z \backslash (ShZ \cup \{z_1, ..., z_n\})) \subseteq r$.*

Then D can be replaced by $D - (r, \otimes)$ and vice versa and we say $(r, \otimes)$ is ***redundant*** *in D.*

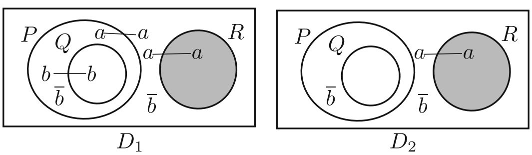

Fig. 11. Redundant i-sequences.

Considering i-sequences, in Fig. 11 two of them are redundant in D_1:

1. The a-sequence with a node in the shaded region tells us that $a \in P' \cap Q' \cap R'$. From this, we can deduce that $a \in P' \cap Q' \cap R'$ or $a \in P \cap Q' \cap R'$, asserted by the other a-sequence, so this second a-sequence is redundant.
2. The presence of the two $\overline{b}$-sequences, together with the shading, allows us to infer that $b \in P \cap Q' \cap R'$ or $b \in P \cap Q \cap R'$, expressed by the b-sequence. Thus, the b-sequence is redundant. Again, we see that the *absence* of the individual b has permitted a reduction in diagram clutter.

Removing the i-sequences from D_1 reduces the clutter score from 11 to 5 in D_2.

Inference Rule 6 (Remove i-Sequence). *Let $D_1 = (L, Z, ShZ, \rho_\otimes, \rho_i, \rho_{\overline{i}})$ be a Venn-i^e diagram and let (r, ι) be an i-sequence in D such that either:*

1. *the region r includes all non-shaded zones: $Z \backslash ShZ \subseteq r$,*
2. *there is a distinct i-sequence, (r', ι), in D such that $r' \backslash ShZ \subseteq r$, or*
3. *there is a set of $\overline{i}$-sequences, say $I = \{(z_1, \iota), ..., (z_n, \iota)\}$ such that $(r \cup \{z_1, ..., z_n\}) \backslash ShZ = Z \backslash ShZ$.*

*Then D can be replaced by $D - (r, \iota)$ and vice versa and we say (r, ι) is **redundant** in D.*

Importantly, all inference rules preserve semantics. In addition, other than the swap rules, applying them never increases diagram clutter. These two properties are captured in Theorem 2.

Theorem 2 (Soundness and Clutter Reduction). *Let D and D' be a Venn-i^e diagrams such that D' is obtained from D by applying one of the inference rules. Then D and D' are semantically equivalent and if the inference rule applied was not a swap rule then the clutter score of D' is at most that of D.*

We are now in a position to show how to minimize clutter in consistent diagrams. Algorithm 1 presents the steps in detail. Referring to Fig. 8, the input to Algorithm 1 is D. Step 1 iteratively removes $\overline{i}$-sequences using inference rule 2, of which D has just one (namely $\overline{a}$), to give D_1. Step 2 iteratively reduces sets of i-sequences using inference rule 4. In this case, the set of a-sequences is reducible and the result is shown in D_2. Taking D_2, step 3 reduces all reducible sequences using inference rule 3; here the result is D_3, where two sequences have altered due to the presence of shading. Step 4 proceeds to remove redundant sequences using inference rule 5, resulting in D_4. Lastly, step 5 inspects the i-sequences to see whether clutter is reduced by swapping them for $\overline{i}$-sequences. In this case, it is beneficial to swap b for two $\overline{b}$s: the b-sequence contributes 5 to the clutter score, whereas the (swapped) $\overline{b}$-sequences in D_{min} contribute just 2. However, the a-sequence contributes only 1 to the clutter score of D_4, so is retained, not swapped. D_{min} is the output from Algorithm 1. Lastly, we note that minimally cluttered diagrams are not, in general, unique. It should be clear from the last step of Algorithm 1 that it is sometimes possible to swap sequences without altering the clutter score.

Theorem 3 (Clutter Minimization). *Let D be a consistent Venn-i^e diagram and let D_{min} be the result of applying Algorithm 1 to D. Then D and D_{min} are semantically equivalent and D_{min} is minimally cluttered.*

The proof can be found online [2].

Algorithm 1. Clutter Minimization

Input: a consistent diagram $D = (L, Z, ShZ, \rho_{\otimes}, \rho_i, \rho_{\bar{i}})$.
Minimise the clutter in D using the following steps:

1. Iteratively swap all $\bar{i}$-sequences, (z, ι), in D for an i-sequence using rule 2. Call the resulting diagram D_1.
2. Iteratively reduce all reducible sets of sequences in D_1 using rule 4 until no reducible sets of sequences remain. Call the resulting diagram D_2.
3. Iteratively reduce all reducible sequences in D_2 using rule 3 until no reducible sequences remain. Call the resulting diagram D_3.
4. Iteratively remove redundant sequences from D_3 using rules 5 and 6 until no redundant sequences remain. Call the resulting diagram D_4.
5. Swap all i-sequences, (r, ι), in D_4, where

$$|Z \backslash (ShZ \cup r)| < 2|r| - 1,$$

for $\bar{i}$-sequences using rule 1. Call the resulting diagram D_{min}.

Output: D_{min}.

---

7 Cognitive Implications

As we have seen, it is possible to reduce clutter in a diagram by removing sequences, reducing them and swapping between i-sequences and $\bar{i}$-sequences. Whilst earlier research into diagram clutter has established that increasing clutter levels correlates with decreased task performance, it is unclear whether and when this remains true for Venn-i^e diagrams. We conjecture that the impact of clutter on task performance will be task dependent.

For instance, consider the semantically equivalent diagrams in Fig. 8 and suppose that we are asked to determine the set in which the individual a lies. We conjecture that this task is easier to perform by studying D_{min} than by studying D. This is because a is more salient in D_{min}, due to the reduced amount of syntax present: this could make it quicker to identify the location of a. Thus, for this task, it could be that D_{min} promotes improved task performance.

Suppose now that our task is to determine whether b is not in P. D_{min} explicitly represents this information using absence (i.e. $\bar{b}$), whereas it must be deduced from D: identify the location of b and deduce that b is not in P. Here, we conjecture that the use of absence has *directly* aided performance. Indeed, there are other tasks for which neither D nor D_{min} are potentially 'optimal'. For example, suppose we wish to determine the set in which b lies. Perhaps the best representation of this information is D_4, which includes a three-node b-sequence (by contrast, D, D_1, D_2 and D_3 are more cluttered). From D_4, we can read off the fact that either b is in just R, b is in just Q, or b is in none of P, Q, and R.

In summary, these examples demonstrate that the diagram that best supports task performance need not be that which is minimally cluttered. There is likely to be trade-off between clutter and directly representing statements of interest, using either absence or presence information. There is clearly an interplay between diagram clutter, the use of syntax to represent presence versus absence and task performance. It is an interesting avenue of future work to explore, empirically, the relationship between diagram clutter and the directness of information representation with respect to task performance.

8 Discussion and Conclusion

In this paper we have explored the potential cognitive benefits of directly representing the absence of individuals in Euler diagram logics. Through identifying sound inference rules, and conditions under which diagrams are inconsistent, we have been able to algorithmically produce minimally cluttered Venn-i^e diagrams. As a consequence, it is possible to represent information about sets and their elements in a minimally cluttered way. The inspiration for this research was derived from related work on Euler diagrams which established that increasing levels of clutter diminished task performance. Our discussion above highlights that the case for reducing clutter in Venn-i^e diagrams, as a way of improving task performance, is less clear cut. Our results lay an essential foundation for empirically evaluating the impact of clutter from this perspective.

As well as empirical research, future work also includes considering clutter and absence in non-classical logics. In our interpretation of Venn-i^e, $\overline{a}$ is syntactic sugar of which we have made use for its practical ability to reduce clutter. There are two other (non-classical) interpretations of Venn-i^e, explored in Choudhury and Chakrabory's work [3]:

1. The absence of a in P does not necessarily imply a is in the complement of P, and
2. The universe is open, so the complement of P does not exist.

In our opinion, the two alternative interpretations are interesting from the point of view of the philosophy and logic of diagrams, and we plan to make them the subject of future work. In the first interpretation, we can represent recursively enumerable sets, which have many important applications in computer science and elsewhere. In the second interpretation since P' does not exist it is also the case that $\overline{a} \in P$ does not imply $a \in P'$. The implications of diagrammatic reasoning with an open universe is an interesting and open topic. Lastly, the use of absence could be incorporated into other Euler-diagram-based logics, such as spider diagrams [9], Euler/Venn diagrams [19], constraint diagrams [7] and concept diagrams [10].

Acknowledgement. This research was supported by EPSRC grant EP/M011763/1.

References

1. Alqadah, M., Stapleton, G., Howse, J., Chapman, P.: Evaluating the impact of clutter in Euler diagrams. In: Dwyer, T., Purchase, H., Delaney, A. (eds.) Diagrams 2014. LNCS, vol. 8578, pp. 108–122. Springer, Heidelberg (2014)
2. Burton, J., Chakraborty, M., Choudhury, L., Stapleton, G.: Technical report: minimizing clutter using absence in Venn-i^e. http://readableproofs.org/diagrams-2016-constants-and-absence-technical-report. Accessed March 2016
3. Choudhury, L., Chakraborty, M.: On representing open universe. Stud. Logic **5**(11), 96–112 (2012)
4. Choudhury, L., Chakraborty, M.: Singular propositions, negation and the square of opposition. Log. Univers. **10**(2), 215–231 (2016)
5. Choudhury, L., Chakraborty, M.K.: On Extending Venn Diagram by Augmenting Names of Individuals. In: Blackwell, Alan F., Marriott, Kim, Shimojima, Atsushi (eds.) Diagrams 2004. LNCS (LNAI), vol. 2980, pp. 142–146. Springer, Heidelberg (2004)
6. Collins, C., Penn, G., Sheelagh, M., Carpendale, T.: Bubble sets: revealing set relations with isocontours over existing visualizations. IEEE Trans. Vis. Comput. Graph. **15**(6), 1009–1016 (2009)
7. Fish, A., Flower, J., Howse, J.: The semantics of augmented constraint diagrams. J. Vis. Lang. Comput. **16**, 541–573 (2005)
8. Horn, L.: A Natural History of Negation, 1st edn. CSLI Lecture Notes. Center for the Study of Language and Information, Stanford (2001)
9. Howse, J., Stapleton, G., Taylor, J.: Spider diagrams. LMS J. Comput. Math. **8**, 145–194 (2005)
10. Howse, J., Stapleton, G., Taylor, K., Chapman, P.: Visualizing ontologies: a case study. In: Aroyo, L., Welty, C., Alani, H., Taylor, J., Bernstein, A., Kagal, L., Noy, N., Blomqvist, E. (eds.) ISWC 2011, Part I. LNCS, vol. 7031, pp. 257–272. Springer, Heidelberg (2011)
11. John, C., Fish, A., Howse, J., Taylor, J.: Exploring the notion of 'clutter' in Euler diagrams. In: Barker-Plummer, D., Cox, R., Swoboda, N. (eds.) Diagrams 2006. LNCS (LNAI), vol. 4045, pp. 267–282. Springer, Heidelberg (2006)
12. Meulemans, W., Riche, N.H., Speckmann, B., Alper, B., Dwyer, T.: Kelpfusion: a hybrid set visualization technique. IEEE Trans. Vis. Comput. Graph. **19**(11), 1846–1858 (2013)

13. Peirce, C.: Collected Papers, vol. 4. Harvard University Press, Massachusetts (1933)
14. Riche, N., Dwyer, T.: Untangling Euler diagrams. IEEE Trans. Vis. Comput. Graph. **16**(6), 1090–1099 (2010)
15. Shimojima, A.: Inferential and expressive capacities of graphical representations: survey and some generalizations. In: Blackwell, A.F., Marriott, K., Shimojima, A. (eds.) Diagrammatic Representation and Inference. LNCS, vol. 2980, pp. 18–21. Springer, Heidelberg (2004)
16. Shin, S.-J.: The Logical Status of Diagrams. Cambridge University Press, Cambridge (1994)
17. Shin, S.-J.: The Iconic Logic of Peirce's Graphs. Bradford Book, Bradford (2002)
18. Stapleton, G., Taylor, J., Howse, J., Thompson, S.: The expressiveness of spider diagrams augmented with constants. J. Vis. Lang. Comput. **20**, 30–49 (2009)
19. Swoboda, N., Allwein, G.: Heterogeneous reasoning with Euler/Venn diagrams containing namedconstants and FOL. In: Euler Diagrams 2004. ENTCS, vol. 134. Elsevier (2005)