Effective Representation of Information: Generalizing Free Rides

Gem Stapleton[1(✉)], Mateja Jamnik[2], and Atsushi Shimojima[3]

[1] University of Brighton, Brighton, UK
g.e.stapleton@brighton.ac.uk
[2] University of Cambridge, Cambridge, UK
mateja.jamnik@cl.cam.ac.uk
[3] Doshisha University, Kyoto, Japan
ashimoji@mail.doshisha.ac.jp

Abstract. In order to effectively communicate information, the choice of representation is important. Ideally, a representation will aid readers in making desired inferences. In this poster, we introduce the theory of *observation*: what it means for one statement to be observable from another. Using observability, we sketch a characterization of the *observational advantages* of one representation over another. By considering observational advantages, people will be able to make better informed choices of representations. To demonstrate the benefit of observation and observational advantages, we apply these concepts to set theory and Euler diagrams. We show that Euler diagrams have significant observational advantages over set theory. This formally justifies Larkin and Simon's claim that "a diagram is (sometimes) worth ten thousand words".

Keywords: Knowledge representation · Observation · Free rides · Inference

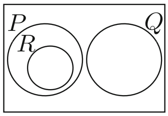
Introduction. When we want to share and understand information, we need to represent it in some notation. There is a plethora of notations available to us for this purpose. This poster, which summarizes [3], is concerned with the relative advantages of one choice of representation of information over another. Many factors can contribute to such advantages. For instance, graphical features, such as the way in which colour is used, and visual clutter (or lack thereof) can impact the ease with which information can be extracted from a representation. The particular focus of this poster is on what we call *observational advantages*.

Observation. It is *advantageous* if a representation of information allows us to simply observe other statements of interest to be true. By contrast, if we cannot observe the statement – yet it does indeed follow from the given representation – then this is a *disadvantage* of that representation. If one representation of information, r_1, has such an advantage and another, r_2, has this as a disadvantage then r_1 has an *observational advantage* over r_2. As a simple example, suppose

M. Jamnik et al. (Eds.): Diagrams 2016, LNAI 9781, pp. 296–299, 2016.
DOI: 10.1007/978-3-319-42333-3_28

$(P \cap Q) = \emptyset$
$R \subseteq P$

(1) Set Theory



(2) Euler Diagram

Fig. 1. Multiple choices of representation.

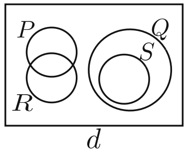


Fig. 2. Free rides.

we wish to represent these two facts about three sets, P, Q and R: nothing is in both P and Q, and everything in R is also in P. There are many notations that can express this information: two examples are illustrated in Fig. 1.

Each of the sentential statements has a single *meaning-carrying relationship*. By meaning-carrying relationship, we mean a relation on the syntax of the statement that carries semantics and evaluates to either 'true' or 'false'. The first statement asserts that the intersection of the two sets is empty. The meaning-carrying relationship in (1) is that $(P \cap Q)$ and $\emptyset$ are written either side of $=$. The statement $(P \cap Q) = \emptyset$ evaluates to either true or false, depending on the interpretation of P and Q as sets. The second statement in (1) describes a subset relationship; here, the meaning carrier in (1) is that R is written on the left of $\subseteq$ and P is written on the right.

The diagrammatic representation has *many* meaning-carrying relationships. The Euler diagram uses non-overlapping curves to express the disjointness of P and Q and, similarly, curve containment to assert that R is a subset of P. Here, two meaning-carrying relationships (namely, disjointness and containment) are exploited to convey the desired information. As a consequence of the way in which Euler diagrams are formed, *additional* meaning-carrying relationships are evident. Most obviously, the non-overlapping relationship between Q and R is a meaning carrier. Thus, from the Euler diagram we can *observe* the statement that Q and R are disjoint. By contrast, this statement cannot be observed from (1) but must be *inferred* from the statements given. This is an example of an *observational advantage* of the Euler diagram over the sentential representation.

Observation can be applied to an Euler diagram to produce another statement, be it a diagram or a set-relation. It can also be applied to a set-relation to produce another set-relation or an Euler diagram. Thus, observation from a single statement, σ, is a binary relationship between σ and another statement, σ_o, denoted $\sigma \rightsquigarrow \sigma_o$, which ensures the following properties hold:

1. some of the meaning-carrying relationships holding in σ hold in σ_o, and
2. σ_o supports just enough relationships to express the meanings carried by the selected relationships in σ and nothing stronger.

Observational Advantages. The new concept of an *observational advantage* generalizes free rides introduced previously by Shimojima [2]. Our definition of an observational advantage requires three key notions to be defined: *semantic entailment*, *semantic equivalence*, and what it means for a statement to be *observable* from a set of statements. The original idea of a free ride assumes a semantics-preserving translation from one notation, N_1, into another notation, N_2, such that the translation ensures the original statements are observable from the resulting statements. We can explain free rides in detail by appealing to our chosen case study: set theory and Euler diagrams. Suppose we have a finite set of set-relations, $\mathcal{S}$, where a set-relation is a statement that asserts either set equality or a subset relationship. Further, suppose that we then identify a semantically equivalent, finite set of Euler diagrams, $\mathcal{D}$, such that each statement, s, in $\mathcal{S}$ is observable from a diagram, d, in $\mathcal{D}$; we can view $\mathcal{D}$ as being a translation of $\mathcal{S}$. Then the set-relations that are observable from the diagrams in $\mathcal{D}$ but not from $\mathcal{S}$ are *free rides* from $\mathcal{D}$ given $\mathcal{S}$.

For example, take $\mathcal{S} = \{(P \cap Q) = \emptyset, (R \cap Q) = \emptyset, S \subseteq Q\}$, which contains three set-relations, and $\mathcal{D} = \{d\}$, where d is in Fig. 2. The free rides from $\mathcal{D}$ given $\mathcal{S}$ are the set-relations that one can observe to be true from $\mathcal{D}$ but which need to be inferred, not simply observed, from $\mathcal{S}$. For instance, we can *observe* both $(P \cap S) = \emptyset$ and $(R \cap S) = \emptyset$ from d but both of these must be *inferred* from $\mathcal{S}$; in the former case, $(P \cap S) = \emptyset$ can be inferred from $(P \cap Q) = \emptyset$ and $S \subseteq Q$. By contrast, whilst the set-relation $(P \cap Q) = \emptyset$ can be observed from $\mathcal{D}$ it can also be observed from $\mathcal{S}$, so it is not a free ride from $\mathcal{D}$. Free rides are examples of what we call *observational advantages* of the Euler diagram over the original set theory representation of information. The difference between observational advantages and free rides is that observational advantages do not require the set $\mathcal{S}$ to contain only statements observable from $\mathcal{D}$.

Set Theory and Euler Diagrams. By applying our theory of observation and observational advantages to set theory and Euler diagrams, we can establish:

1. Given a finite set of set-relations, $\mathcal{S}$, no other set-relations can be observed from $\mathcal{S}$; thus, $\mathcal{S}$ is *observationally devoid*.
2. Given an Euler diagram, $d_{\mathcal{S}}$, constructed from $\mathcal{S}$, every set-relation that follows from $\mathcal{S}$ can be observed from $d_{\mathcal{S}}$; thus, $d_{\mathcal{S}}$ is *observationally complete*.

These two characterizations of what can (or cannot) be observed allow us to understand that $d_{\mathcal{S}}$ is a significantly more efficacious representation of information than $\mathcal{S}$: it has *maximal observational advantage* over $\mathcal{S}$. Thus, from a purely inferential point of view, using $d_{\mathcal{S}}$ is desirable: $d_{\mathcal{S}}$ makes informational content readily available, in the sense of observability, to end-users. As there are infinitely many set-relations that are semantically entailed by $\mathcal{S}$, the benefits of Euler diagrams over set theory are numerous.

Conclusion. In our view, the result introduced here captures the kernel in which diagrams facilitate our inference and thus excel over sentential representations. Putting this in a larger perspective, we expect to gain a fuller understanding of the relative advantages of one choice of representation over another. Linking back to the insight that a diagram is sometimes worth 10,000 words [1], our formal theory of observation and observational advantage has allowed us to prove that a diagram is sometimes worth *infinitely many* set-relations.

References

1. Larkin, J., Simon, H.: Why a diagram is (sometimes) worth ten thousand words. Cogn. Sci. **11**, 65–99 (1987)
2. Shimojima, A.: Semantic Properties of Diagrams and Their Cognitive Potentials. CSLI Publications, Stanford (2015)
3. Stapleton, G., Jamnik, M., Shimojima, A.: What makes an effective representation of information: a formal account of inferential advantages. J. Log. Lang. Inf. (2015, Submitted)