JOURNAL PRE-PROOF

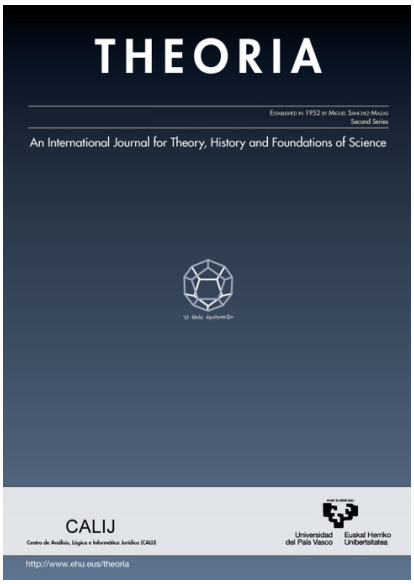
Quantifying information in structural representations

Stephen Mann

DOI: 10.1387/theoria.25212

Received: 03/11/2023

Final version: 15/10/2024



Quantifying information in structural representations

(Cuantificación de información en representaciones estructurales)

Stephen Francis Mann*

Max Planck Institute

Abstract: The goal of this paper is to show that the information carried by a structural representation can be decomposed into the information carried by its component parts. In particular, the relations between the components of a structural representation carry quantifiable information about the relations between components of their signifieds. It follows that the information carried by cognitive structural representations, including cognitive maps, can in principle be quantified and decomposed. This is perhaps surprising given that the formal tools of communication theory have typically only been applied to simpler representation-like states without significant structure, such as detectors or indicators. In the final section I consider using computational complexity theory to capture the processing advantages afforded by structural representation.

Keywords: representation, structural representation, information theory, mutual information, cognitive science, computational complexity theory.

Resumen: El objetivo de este artículo es demostrar que la información contenida en una representación estructural puede descomponerse en la información contenida en sus partes componentes. En particular, las relaciones entre los componentes de una representación estructural contienen información cuantificable sobre las relaciones entre los componentes de sus significados. De ello se deduce que la información contenida en las representaciones estructurales cognitivas, incluidos los mapas cognitivos, puede en principio cuantificarse y descomponerse. Esto es quizás sorprendente dado que las herramientas formales de la teoría de la comunicación generalmente solo se han aplicado a estados más simples similares a representaciones sin estructura significativa, como detectores o indicadores. En la

*Correspondence to: Stephen Francis Mann. Department of Linguistic and Cultural Evolution, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, Leipzig 04103 Germany – stephenfmann@gmail.com – ORCID: 0000-0002-4136-8595

sección final, considero el uso de la teoría de la complejidad computacional para capturar las ventajas de procesamiento que ofrece la representación estructural.

Palabras clave: representación, representación estructural, teoría de la información, información mutua, ciencia cognitiva, teoría de la complejidad computacional.

Short summary: I demonstrate how to quantify the informational content of structural representations. In particular, I show that the relations borne between components of a structural representation carry quantifiable information about the relations borne between components of the representation's signified. The use of information theory to capture aspects of cognitive processing is therefore broader than is typically assumed.

1. Introduction

The goal of this paper is to show that the information carried by a structural representation can be decomposed into the information carried by its component parts. In particular, the relations between the components of a structural representation carry quantifiable information about the relations between components of their signifieds. It follows that the information carried by cognitive structural representations, including cognitive maps, can in principle be quantified and decomposed. This is perhaps surprising given that the formal tools of communication theory have typically only been applied to simpler representation-like states without significant structure, such as detectors or indicators.

Towards the end of the paper I discuss a puzzle raised by the above argument. If structural representations carry information just as unstructured representations do, then the purported benefits of structural representation cannot be explained on the basis of information-carrying alone. I raise the question of what does explain the benefits of structural representation, canvassing intuitive arguments given informally by philosophers of cognitive science. I then ask whether formal results from computational complexity theory can support these informal arguments, concluding that they cannot yet be said to do so. If we want to draw on formal work to support informal theorising about representation (a methodological stance I admit might not be shared by all), we have to apply computational complexity theory in a more subtle way than has so far been attempted in cognitive science.

1.1 Background: indicator states and structural representations

I hope to contribute to our understanding of a distinction, common in the philosophy of cognitive science, between indicator states and structural representations. Indicator states are simple signals, without significant structure, that are typically taken to co-occur with their signifieds. Examples include sensory registration (Burge, 2010, p. 315) and signals produced by receptors (Ramsey, 2007, ch. 4). Structural representations are more complex cognitive states, typically detached and persistent over time, that bear a structure-preserving resemblance relation to their signifieds. The canonical example is a cognitive map (Shea, 2018, §5.2). Indicator states are often discussed in terms of informational measures like mutual information, while discussions of structural representation typically omit these terms entirely. If my argument holds water, we would have good grounds for thinking informational measures apply to structural representations too, not just indicator states.

1.2 Motivation: the explanatory role of informational measures

Philosophers of mind and cognitive science have long considered the relationship between informational measures and semantic content (Dretske, 1981; Gallistel, 2020; Shea, 2018). Some have tried to characterise the explanatory role of mutual information in a way that cashes out informal philosophical theorising about the content of signs. Their hope is that the established explanatory role of mutual information can shed light on a problem about the explanatory role of semantic content. That problem is as follows: behaviour triggered by a sign can only be sensitive to the properties of the sign vehicle, not any properties of the signified. If content is supposed to be something over and above the vehicular properties of the sign, it seems as though it cannot play a role in the explanation of behaviour. One

prominent resolution to this problem is to appeal to content to explain success, then to explain behaviour derivatively (Shea, 2018, §2.3). And for those who wish to tie philosophical theorising to formal approaches in science, the explanatory role of mutual information can help make this link. Mutual information tells us how an agent can condition her behaviour on a sign and act optimally across a range of situations. The information in a sign explains behaviour because it explains why the agent has learned *that conditionalisation* in the first place. In terms famously introduced by Dretske (1988, §2), the information in a sign is a structuring cause that explains why this particular sign vehicle is a triggering cause for this particular behaviour.[1]

Since the role of mutual information in a formal context is similar to the role of semantic content in an informal context, it is tempting to try and link the two. And although there is no consensus on how this link should be forged, demonstrating the applicability of informational measures to structural representations widens the scope of any theory that connects such measures to semantic content. Gallistel (2020) points out that the conceptual framework provided by applying information theory to problems of communication and computation is broader than most philosophers admit. This paper lends support to his point by demonstrating that the applicability of mutual information and related measures is wider than has often been appreciated.

1.3 Structure of the paper

Section 2 describes the standard way of applying information-theoretic measures to unstructured signs. Section 3 suggests a way to apply those measures to features of structured signs, particularly to the relations between sign components. Section 4 discusses the possibility of using computational complexity theory to understand the benefits of structural representation. Section 5 concludes.

2. Information as correlation

2.1 Background

The use of informational measures to quantify correlations started with the introduction of **mutual information**.[2] Mutual information (originally called rate of transmission) was introduced by Shannon (1948, p. 407) in the foundational text of communication theory. It

[1] An anonymous reviewer highlighted recent evidence, presented by Favela and Machery (2023), that neuroscientists and psychologists don't attribute representational status on the basis of function. This result is concerning for philosophers who want to define cognitive representations in terms of function or who want to characterise the explanatory role of representation in terms of successful behaviour. I'm following the mainstream view and can't engage fully with those survey results. I'll just make two short points in defence of my approach: first, Richmond (2023) argues that Favela & Machery's conclusions are unwarranted on the basis of their study, so we shouldn't be quick to jettison a functional account of representation; second, functional accounts like that given by Shea (2018) are based on a careful analysis of the explanatory role of representation in actual cognitive science case studies, which arguably provides a more accurate picture of what the concept of representation is doing in the discipline than does a survey of its practitioners.

[2] Fisher information was defined earlier, but is not nowadays mentioned in philosophical discussions of correlation and semantic content. It's questionable whether the relationship Fisher information quantifies can reasonably be called correlation (it's a relationship between a random variable and a parameter of a statistical model, rather than between two random variables), so I've ignored it in this paper.

was used to quantify how much could be learned about a signal transmitted through a channel by observing the signal received. Noise in the channel can corrupt a signal, leading the received signal to differ from the signal transmitted. Since noise is modelled as a probabilistic change to the signal governed by known statistical parameters, the received signal provides probabilistic evidence about the transmitted signal. Quantifying the amount of probabilistic evidence was useful for the goals of communication theorists, and mutual information turned out to be a particularly appropriate measure.

It was soon recognised that mutual information applies beyond the context of signals transmitted through channels. It is a general measure of how much can be learned about an unobserved process from a correlated observed process. Its mathematical form, which we will introduce in a moment, lends itself to a wide variety of applications. Sciences in which mutual information is nowadays used to quantify correlations include behavioural ecology (Haldane & Spurway, 1954), cosmology (Pandey & Sarkar, 2017), linguistics (Hunston, 2002), molecular biology (Mehta et al., 2009), and neuroscience (Rathkopf, 2017). Its generality can be illustrated with an example.

2.2 Example 1: coin and light

Suppose a fair coin is tossed and lands heads or tails. At the same time there is a light which can be green or red. Suppose that when the coin lands heads, the light always shows green, and when the coin lands tails, the light always shows red. Intuitively one can learn which way the coin landed by observing the colour of the light. There is a very strong correlation because the two sets of states perfectly correspond to each other.

How can we quantify this correlation? Although there are various statistical techniques to measure correlations, many of them require numeric data. For example, measuring the covariance or Pearson correlation coefficient between the coin and the light would require assigning numeric values like 0 and 1 to the different coin faces and light colours. While this might be possible and interpretable when there are just two states, it's not clear how to assign numeric values in a non-arbitrary way when an event has more than two non-numeric outcomes. If the light could be green, red or yellow, there is no obvious way to assign numbers to those values in order to measure covariance in an interpretable way.[3]

In contrast to traditional statistical measures, mutual information is defined in terms of a joint probability distribution across the different possible states of each variable. This does not require that the variables take numeric values, only that their collective probabilities can be defined. I will now give the definition of mutual information.

2.3 Defining mutual information

Mutual information can be understood as a measure of how different the **joint probabilities** of two variables are from how they would be if the variables were not statistically associated. Continuing our example, we can construct a joint probability distribution describing the relationship between the coin and the light (table 1). Mutual information uses the joint distribution to measure the strength of correlation between coin and light.

[3] One non-obvious technique employed by statisticians is dummy coding. The three colours are treated as three different binary variables, so that Y_1 is the event 'the light is green', with values of 1 and 0 standing for 'yes' and 'no', and Y_2 and Y_3 standing for red and yellow lights accordingly. While these manipulations are possible, they are a rather unwieldy way of squeezing non-numeric data into numeric form.

	green	red
heads	$\frac{1}{2}$	0
tails	0	$\frac{1}{2}$

Table 1: A joint probability distribution describing the statistical relationship between a fair coin and a light that can be green or red. Half the time the coin lands heads and the light is green. Half the time the coin lands tails and the light is red. The other combinations of coin face and light colour never occur.

Letting the coin be denoted by X and the light by Y, the actual joint probabilities are $p(X, Y)$ while the distribution one would see if they were uncorrelated is the product of unconditional probabilities $p(X)p(Y)$.[4] Mutual information is constructed by taking the log ratio for each value of $x \in X$ and $y \in Y$, $\log \frac{p(x,y)}{p(x)p(y)}$, and summing over all possible situations,[5] weighted by the actual probability of each situation $p(x, y)$:

$$I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

When X and Y are statistically independent the mutual information is zero, its minimum possible value. That's because statistical independence entails that the joint probability is equal to the product of unconditional probabilities for every x and y: $p(x, y) = p(x)p(y)$. The ratio inside the logarithm is therefore 1 for every x and every y, which means that the logarithm is 0 for every pair, which means the overall sum is zero. Mutual information takes its maximum value when events are perfectly correlated, as in table 1. This maximum is in general unbounded, and is determined by the number of events in X and Y, their unconditional distributions and their conditional distributions with respect to each other. With a minimum value of zero and a maximum that is determined by the probabilistic relationship between X and Y, mutual information is a nice way to formalise the intuitive concept of how strongly correlated two events are. So one way to interpret mutual information is as an answer to the question: how strongly correlated are X and Y?

We can now calculate the mutual information between the coin and the light. Taking the logarithm to base 2, the mutual information between the coin and the light is:

[4] Throughout I'll use upper-case letters inside $p(.)$ to denote probability distributions and lower-case letters inside $p(.)$ to denote specific probabilities.

[5] Here and throughout I am referring to the discrete form of mutual information which is defined in terms of probability distributions. As an anonymous reviewer rightly pointed out, there is a continuous version which is defined in terms of probability densities, and many of the situations in which we might want to apply mutual information would be more suited to the continuous form. Since I'm aiming for simplicity in order to make a prima facie case for treating structural representations in terms of information, I will continue to just talk about the discrete form. Future work should investigate the consequences of adopting the continuous form.

$$
\begin{aligned}
I(X;Y) &= \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \\
&= \frac{1}{2} \log \frac{\frac{1}{2}}{\frac{1}{2}\cdot\frac{1}{2}} + 0 + 0 + \frac{1}{2} \log \frac{\frac{1}{2}}{\frac{1}{2}\cdot\frac{1}{2}} \\
&= \frac{1}{2} \log 2 + \frac{1}{2} \log 2 \\
&= 1 \text{ bit}
\end{aligned}
$$

It is commonplace to report this result with a locution like 'the light carries one bit of information about the coin'. This means that the light, as an event or process that can take one of two states, carries information about a coin flip, which is also an event or process that can take one of two states. Mutual information is measured over both event spaces, not over individual outcomes. In order to determine informational properties of individual outcomes, we must turn to the decomposition of mutual information into its component parts.

2.4 Decomposing mutual information

Philosophical discussions of information often advert to specific outcomes 'changing the probabilities' of other outcomes. The probability of the coin landing heads (or having landed heads) is in a loose sense changed by the occurrence of a green light, because the unconditional probability of heads is $\frac{1}{2}$ while the conditional probability of heads given a green light is 1. We can take the log ratio between these to get what's known as **pointwise mutual information**: $\log \frac{p(x|y)}{p(x)} = \log \frac{1}{\frac{1}{2}} = \log 2 = 1$ bit. This is a measure between the specific outcomes $x =$ heads and $y =$ green. We can build up to a quantity carried by the green light across all situations X by finding the weighted sum of its pointwise mutual information with respect to both coin outcomes (this relationship is described by Skyrms, 2010, §3). The resulting measure is called **relative entropy** and is denoted by D:

$$
D(p(X|Y=y)||p(X)) = \sum_{x} p(x|y) \log \frac{p(x|y)}{p(x)}
$$

Relative entropy measures the information carried by a single event $y \in Y$ (say, the green light) about all coin outcomes X. The relative entropy of the green light with respect to the coin is:

$$
\begin{aligned}
D(p(X|Y=\text{green})||p(X)) &= \sum_{x} p(x|\text{green}) \log \frac{p(x|\text{green})}{p(x)} \\
&= 1. \log \frac{1}{\frac{1}{2}} + 0 \\
&= \log 2 = 1 \text{ bit}
\end{aligned}
$$

The green light carries an average of 1 bit about coin outcomes because it carries 1 bit when the coin lands heads and it doesn't occur when the coin lands tails.[6]

We reach mutual information by taking the weighted sum of relative entropy for both light colours:

$$
\begin{aligned}
I(X;Y) &= \sum_y p(y) D(p(X|Y=y)||p(X)) \\
&= \sum_y p(y) \sum_x p(x|y) \log \frac{p(x|y)}{p(x)} && \text{(definition of } D\text{)} \\
&= \sum_{x,y} p(x,y) \log \frac{p(x|y)}{p(x)} && \text{(defn. joint probability)} \\
&= \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} && \text{(defn. conditional probability)}
\end{aligned}
$$

As we saw, the mutual information between coin and light is 1 bit. It might seem strange that in our example all three quantities take the value 1 bit. It's clear why that happens when you note that relative entropy and mutual information are averages, not totals summed across all possible states. The green light carries 1 bit across both coin face situations, because it carries exactly 1 bit about heads and it never occurs when the coin lands tails. Since both the green and red lights carry 1 bit about the coin, and each has a 50% chance of occurring, the light as a system carries an average of $\frac{1}{2}.1 + \frac{1}{2}.1 = 1$ bit. Of course, if you ran multiple trials the information would start to stack up: the light is conveying 1 bit per trial, so over N trials it would convey a total of N bits. This is a sum rather than an average. Mutual information is giving you the per-trial expected amount of information.

To reiterate, these three measures relate different things or sets of things:

- The amount of information a particular sign y carries about a particular signified x is their **pointwise mutual information**.
- The amount of information a particular sign y carries about a set of signifieds X is their **relative entropy**.
- The amount of information a set of signs Y carries about a set of signifieds X is their **mutual information**.

In the previous subsection we interpreted mutual information as a measure of the correlation between coin and light. Given its decomposition into component terms, another way to interpret mutual information is as an answer to the question: what is the average amount of

[6] Relative entropy is also called Kullback-Leibler divergence and has been used extensively in Bayesian cognitive science, machine learning, and related areas (Itti & Baldi, 2009, p. 1297).

pointwise mutual information carried by each sign about each signified?[7]

In the next section I argue that this decomposition extends further. The components of structured signs can be attributed quantities that describe how much information they carry about the component elements of the signified. The most interesting application is to cognitive maps, but I will start with examples of simpler structured signs.

3. Measuring information in structured signs

Structured signs are those for which relations between sign components correspond to relations between components of world affairs that are the referents of those components. This rather laboured definition will become much clearer when we see a few examples. First I should address the question of what relations are, which is a rather significant topic in ontology (MacBride, 2020). I'm taking a very simple realist approach, of the kind described by Shea (2018, p. 112): "On the thin notion of relation, any set of n-tuples corresponds to a relation (an n-place relation)." Shea goes to great lengths to show how to cut down this huge class of relations to pick out just those that constitute content in structural representations. Fortunately, my task is much simpler: I am not aiming here to give an account of content, but to describe informational relationships between signs and signifieds – including informational relationships between relations within signs and relations within signifieds. All I require is that there are such relations, which many accounts of structural correspondence assume and which the several examples in this section demonstrate. I will not offer a principled way to individuate signs, signifieds and relations, instead relying on examples which are hopefully intuitively acceptable and which motivate the analysis.

3.1 Example 2: the taller-than relation

Let's start with a very simple system. Consider a structured sign that depicts two referents and a relation between them. It has eight components, which we'll label with the first eight letters of the Latin alphabet: A, B, C, D, E, F, G, H. Each possible sign is a combination of two of these letters. We can imagine each letter to represent a person, and each whole sign to be representing one person being taller than another. So the sign AB means 'A is taller than B', which is an asymmetric and non-reflexive relation. Suppose that whatever is producing these signs only takes into account two people at once (i.e. although it's logically possible that A is taller than B and C is taller than D, on a single 'trial' only one of these scenarios is occurrent). Then there are 56 possible signs in this system (seven signs per starting letter, of which there are eight): AB, AC, AD, AE, AF, AG, AH, BA, BC, BD, … HD, HE, HF, HG. To simplify things, we'll say that no two people are the same height, and everyone has an equal chance of being taller than anyone else.

The job of each letter is to indicate a person, and the job of the ordering of the letters is to indicate a relation between two people. What we would like is a way to break down the

[7] There is another sense of 'decomposition' relating to mutual information that should be distinguished from what I've described in this section. A technique known as partial information decomposition (Gutknecht et al., 2021) enables the information carried by multiple signs about a signified to be split into three components: the *unique* information that each sign carries and no other sign does, the *redundant* information that is carried by more than one sign, and the *synergistic* information that is not carried by any individual sign but arises from their combination. Whether or not there is an interesting relationship between the two senses of decomposition is a question for the future.

information quantity of a sign into the information provided by its component parts. Just as we broke down the mutual information across an entire system into the pointwise mutual information associated with each sign-signified pair, we now ask whether signs that are themselves structured can be broken down into some formal object associated with each component. The good news is that we already know how much information each individual sign contains. There are 56 equiprobable signs. Each signifies that a particular outcome, which starts with a probability of $\frac{1}{56}$, is in fact occurrent, giving it a probability of 1. Each sign therefore has a pointwise mutual information with its particular signified of $\log \frac{1}{\frac{1}{56}} = \log 56$ bits of information. This is the same quantity that a more laborious non-compositional system of 56 distinct signs would have. The established decomposition from mutual information to pointwise mutual information works for compositional signs too, because it applies no matter whether the sign itself is simple or compositional.

How can we break down the information in a compositional sign into the information carried by each of its components? The idea is that a compositional sign carries information in each of its components such that these contribute to the total information in the sign. What will be of particular interest is whether the ordering itself carries information; that is, if the relation between the symbolic components of the sign carries information about the relation between the people in the room. Intuitively, it looks like we can make the case both that it does and it does not:

- *Yes, the ordering carries information*: An AB sign is distinguished from a BA sign only by the ordering. Supposing these signs are equiprobable, then the ordering must carry 1 bit of information, because it's enabling us to distinguish between these two cases.
- *No, the ordering does not carry information*: The first component carries $\log 8$ bits (because there are 8 possibilities for it), the second component carries $\log 7$ bits (because there are only 7 possibilities after the first one has been given). When you add logarithms you multiply the arguments, so the total amount of information carried by the two components is $\log 8 + \log 7 = \log 56$ bits. The sign only carries $\log 56$ bits in total, so there's no extra information for the ordering to carry.

Intuitions conflict, but there is something wrong with the 'No' answer. By saying that the first component carries $\log 8$ bits and the second carries $\log 7$, we are assuming that we already know which is which. Until you know the ordering, you do not know which component was counted first. Any means by which you could get $\log 8 + \log 7$ bits of information must be smuggling in information about the relation. The right way to do the sum is to say that knowing the identity of one of the components (but not its position) carries $\log 4$ bits. That's because each symbol appears in exactly 14 of the signs. Knowing one symbol increases the probabilities of each outcome containing the corresponding person from $\frac{1}{56}$ to $\frac{1}{14}$, which delivers $\log \frac{\frac{1}{14}}{\frac{1}{56}} = \log \frac{56}{14} = \log 4$ bits. Once you know this first symbol, being told the second does indeed impart $\log 7$ bits, even when you don't know the order: there are 14 possible signs and two of them contain this particular second symbol; $\log \frac{14}{2} = \log 7$. Overall, knowing the two symbols, but not the order in which they appear, imparts $\log 4 + \log 7 = \log 28$ bits. But we know that the whole sign carries $\log 56$ bits in total. Therefore, since the whole sign carries $\log 56$ bits and the components carry only $\log 28$ bits, we should conclude that the ordering of the components carries $\log 2$ bits. The total information carried by the sign

can therefore be expressed as $\log 28 + \log 2 = \log 56$ bits as required. The upshot is that the relation between sign components carries 1 bit of information about the relation between the referents of those components.

3.2 Interpreting probabilities

Before moving on to more complex examples it's worth reflecting on the question of how to interpret the probabilities that appear in informational measures. Discussions of information in communication theory typically adopt a subjectivist viewpoint because the information provided by a sign is relative to what you already know. The same is true in the examples we've seen so far. The information provided by whole signs is relative to the existing knowledge of an observer: if you already know the coin landed heads, then a green light doesn't provide you with information. The point extends to structured signs, where knowledge of one of the components impacts the information carried by the others: if you already know there is a B in the sign, then finding out there is an A provides $\log 7$ rather than $\log 4$.

Since informational measures are defined in terms of probabilities their values will differ depending on the interpretation of probability one favours. A subjectivist interpretation might say the green light carries different amounts of information depending on who is observing it, while an objectivist interpretation might say it carries 1 bit regardless of who is observing it. These labels aren't perfect, since Scarantino (2015) gives an 'objective-relative' interpretation that mixes them. A better way to understand these different approaches is to treat them as providing different answers to the reference class problem: given a single outcome, with which other outcomes should it be grouped in order to determine a probability distribution over an event space (Hájek, 2007; Millikan, 2013)? Here I'm using 'objectivist' to mean one who wants to define a reference class without appeal to potential observers (e.g. Skyrms, 2010), and 'subjectivist' to mean one who wants to define a reference class by appeal to potential observers (e.g. Millikan, 2013; Scarantino, 2015). For ease of exposition I'm going to continue in a subjectivist or 'relativist' idiom. Given this approach, signs literally change probabilities because they change observers' knowledge with respect to which those probabilities are defined. The amount by which probabilities are changed, and hence the amount of information in a sign, is relative to what the observer already knows. This kind of approach has been defended by Millikan (2013) and Scarantino (2015).

3.3 Example 3: the nominates-for-President relation

Things get a little more complicated when we include repetitions (AA, BB, ...). Perhaps the sign is now indicating a reflexive relation like 'nominates for President'. Because the statistics describing the presence of different components in the sign have changed, the relation carries a different amount of information. If you were told the signal has two As there is no more information to receive about which signal it is, so no information for the relation to carry. Since there are eight of these double signs, and 64 signs overall, there is a $\frac{1}{8}$ chance that the relation provides no information and a $\frac{7}{8}$ chance that it provides 1 bit as before.

The decomposition of mutual information allows us to formalise these observations into an average amount of information carried by the relation. However, there is a prima facie problem measuring relative entropy and mutual information for the relation. Those measures require a variable y that takes values from event space Y. It's not obvious that the

relation can be described in this way. From one perspective, every sign contains the same relation, namely 'nominates for President'. From another perspective, we can describe these relations differently, such as 'nominates themself' and 'nominates someone else'. But it's not clear whether we can accurately say that the difference between AB and BA is that one contains a different relation from the other. We would have to define the values of the relation as something like 'nominates a person designated by a letter later in the alphabet' and 'nominates a person designated by a letter earlier in the alphabet'. Then in the original case these have an exactly 50/50 chance of occurring, which is another way to capture the fact that the relation carries 1 bit.

In the present case, where repetitions are allowed, there is a new value for the relation: 'nominates themself'. The event space describing the relation could therefore be constructed as ⟨nominates later, nominates earlier, nominates themself⟩. What are the probabilities here? The relation 'nominates themself' only occurs when both symbols are the same, and that only happens on eight out of 64 occasions, or $\frac{1}{8}$. The other two relations occur on exactly half of the remaining occasions, and half of the remaining $\frac{7}{8}$ is $\frac{7}{16}$. Therefore the probabilities associated with this way of carving up the event space of relations are $\langle \frac{7}{16}, \frac{7}{16}, \frac{1}{8} \rangle$.

Before calculating informational measures using this distribution, it is worth considering the problem of how to choose the values of the relation variable. Birch (2014, §6) calls the problem of how to carve up an event space the partition problem: "how are states of the world to be individuated in any principled way, outside the context of simple formal models?" (Birch, 2014, p. 508). The partition problem is particularly pressing when we need to use an event space to define some property we think ought to be specified independently of the interests of modellers (in Birch's case this is semantic content).[8] There are a couple of reasons why I think the problem can be sidestepped in our case. First, it is not clear to me whether attributions of information in cognitive science need to be entirely independent of the interests of scientists modelling cognitive representations. If it turns out they do need to be independent, the partition problem would need more attention; however, theoretical interest in information attribution has not been beset by problems of indeterminacy like semantic content has, so the problem is not as vital. Second, since we are adopting a relativist perspective on probability, it seems appropriate to adopt a relativist perspective on partitioning too. In the present example I chose a simple partitioning for ease of exposition; in general, one observer's ability to learn about the world from a sign might differ from another's. As well as the information quantity differing for them, the actual discriminations they can make might differ too. As a result, it might be possible to appeal to observer capacities to determine the appropriate partitioning for any particular sign. This is only the briefest sketch of a way to avoid the partition problem. A thorough analysis would take a whole paper.

Getting back to the matter at hand, the relative entropy and mutual information can be calculated using this distribution for Y. Consider first the relative entropy associated with the relation 'nominates themself'. This describes how much information you gain about who has nominated whom when you learn that someone has nominated themself. Intuitively it should capture the fact that we have gone from 64 possibilities to eight, thus it should be $\log \frac{64}{8} = \log 8 = 3$ bits. The full relative entropy is:

[8] The partition problem discussed here is not to be confused with a mathematical problem of the same name, which has interesting properties from the perspective of computational complexity theory. Although I will raise issues relating to that theory below, the mathematical partition problem has no bearing on the discussion.

$$
\begin{aligned}
D(p(X|Y=y)||p(X)) &= \sum_x p(x|y) \log \frac{p(x|y)}{p(x)} \\
&= \frac{1}{8} \log \frac{\frac{1}{8}}{\frac{1}{64}} + 0 + 0 + ... + \frac{1}{8} \log \frac{\frac{1}{8}}{\frac{1}{64}} \\
&= 8. \left(\frac{1}{8} \log \frac{\frac{1}{8}}{\frac{1}{64}} \right) \\
&= \log \frac{64}{8} = 3 \text{ bits.}
\end{aligned}
$$

So the relation 'nominates themself' carries 3 bits. Similar calculations show that the other two relations carry $\log \frac{16}{7}$ (slightly more than 1) bits each. Therefore the mutual information carried by this set of relations about the state of affairs is:

$$
\begin{aligned}
I(X;Y) &= \sum_y p(y) D(p(X|Y=y)||p(X)) \\
&= \frac{7}{16} \log \frac{16}{7} + \frac{7}{16} \log \frac{16}{7} + \frac{1}{8} \log 8 \\
&\approx 1.419 \text{ bits.}
\end{aligned}
$$

Furthermore, different pieces of evidence alter the expected contribution of the relation. Suppose you are told that one of the relata is an A and the other is either an A or a B. Then there are three possible signs: AA, AB, BA. Each of the possible relations appears exactly once in this set of signs. Therefore, learning the relation uniquely picks out one sign from three, which entails that each relation imparts $\log 3$ bits. Each relation has changed the amount of information it is providing. Receiving different pieces of information about the components changes the expected contribution of the relation. This result is consistent with the relativist perspective on which information quantities are calculated with respect to a user's knowledge.[9]

3.4 Example 4: a simple map

Philosophers of cognitive science are often concerned with the rich spatial structure of cognitive maps, so we should apply this analysis to spatial relations. Consider a discrete map with four quadrants, North, South, East and West. Each quadrant can be one of two colours, red or blue. As with all maps, every quadrant bears a spatial relation to every other quadrant. Note that these relations are not necessarily unique: North bears the same relation to East as West does to South (namely, being to the upper-left of).

As before, our intuitions might conflict as to whether spatial relations are carrying quantifiable information. There are $2^4 = 16$ possible maps, so each map carries $\log 16 = 4$

[9] The discussion in this section focuses on the information carried by the structural relation about the whole state of affairs. The information carried by the relation about the relation itself would just be the surprisal of that relation, e.g. the information carried by the 'nominates themself' relation about the actual relation between nominator and nominee is $\log \frac{1}{\frac{1}{8}} = 3$ bits.

bits (assuming equiprobability as always). Each quadrant's colour carries 1 bit (red or blue), for a total of 4 bits. As before, this is misleading. Learning that there is a red quadrant only imparts $\log \frac{16}{15}$ bits, because 15 of the possible maps have at least one red quadrant. Learning that there are two red and two blue quadrants imparts $\log \frac{16}{6}$ bits due to the six maps with this combination of colours. If you are subsequently told of a relation between two of those colours, you will learn further information. So for example, if you are told that one of the red quadrants is to the upper-left of the other red quadrant, that gives you a further $\log \frac{6}{2} = \log 3$ bits, because the reds must either be in the North and East or in the West and South. On the other hand, if you are told that one of the red quadrants is to the upper-left of one of the blue quadrants, that only imparts $\log \frac{6}{3} = 1$ bit, because there are three possible maps (out of the six you've already whittled it down to) consistent with this new evidence.

As in the nominates-for-President case, the spatial relations between components of a map carry different amounts of information depending on the components themselves. If we wanted, we could calculate the average information carried by a relation by summing over all the representations it can participate in, averaged by their probabilities. In some maps, the spatial relations will carry a great deal of information; in others they may carry very little. Imagine the extreme case where a map happens to depict a territory with nothing in it. Then the relations don't tell you anything. If the map has a distance scale, that might tell you something (i.e. how much of the territory is bare). But the relations between points aren't giving you that rich information usually associated with maps.

It might seem just wrong to try and attribute probabilities to a map, as if it were a Shannonian signal selected from a set of possible signals. One might think that in order to attribute probabilities to a representation, it must be selected from a set of available representations. That is after all how the basic Shannon framework is constructed. The source produces one outcome from a set of possible outcomes, each of which has its own probability of being produced. The encoding scheme converts source outcomes into signals. Sometimes this conversion is deterministic but it can be probabilistic (in any case a deterministic encoding is a limiting case of a probabilistic encoding in which all the probabilities are either 0 or 1). The encoding scheme together with the probability distribution over source outcomes determines a joint probability distribution over source outcomes and signals representing them. If the source distribution is $p(X)$ and the encoding is $p(Y|X)$ then the relevant joint distribution is $p(X).p(Y|X) = p(X, Y)$. It is this joint probability distribution that is required for informational measures, in particular the mutual information between signals and source outcomes.[10]

Given this picture an objection to my account runs as follows. Defining informational measures between a map y and its territory x requires that there be a joint probability distribution between their respective event spaces Y and X. Shannon defines joint probability distributions by starting with a source distribution $p(X)$ and multiplying it by the conditional distribution defined by the encoding scheme $p(Y|X)$. An encoding scheme is a process by which signals are chosen probabilistically given a particular source outcome. Maps and territories cannot be attributed a joint distribution in this way, because maps are not chosen probabilistically given a particular territory. We don't have a stack of maps from which we choose the correct one upon observing the territory. Rather, we construct a single map

[10] Even though the original application of mutual information measured the correlation between signal-before-noise and signal-after-noise, contemporary philosophical discussion tends to follow Skyrms's lead and consider the mutual information between signals and source outcomes.

by combining components that refer to aspects of the territory, in such a way that the map components bear relations to each other that reflect the spatial components of their referents in the territory. Therefore, maps and territories can't be attributed informational measures by the same procedure informational measures are typically derived.

Let's assume that one of the premises of this objection is true (and it does seem obviously true): maps are not selected from among a set of possible maps, and cognitive maps certainly aren't. The question then is whether a joint probability distribution can be defined between a map and its territory. It can't be done in the Shannonian way – we don't have a stack of maps from which we choose the correct one upon observing the territory – but can it be done in some other way? Rather than actually describing such a procedure I'm going to offer an *a fortiori* argument that the standard commitments of Bayesian cognitive science require that such a joint distribution can be defined. This brings us close to a wide-ranging and important debate in the philosophy of cognitive science regarding the propriety of the Bayesian approach. Therefore, I'm going to outline those commitments of the Bayesian agenda that I take to support my own argument here, and leave the wider problems for later. I'm hitching my wagon to the possibility of Bayesian cognitive science, and if that program takes a tumble, I'll go down with it.

So, what is it about Bayesian cognitive science that makes me think joint probability distributions can be attributed to maps and their territories? In short, it is that the Bayesian approach always assumes that unobserved events can be represented with prior distributions, and evidence pertaining to unobserved events can always be represented with conditional distributions (commonly called likelihoods). On a Bayesian view, regardless of what the brain is doing at an algorithmic and implementational level, on the computational level it can always be described as using priors and likelihoods to determine posterior distributions over an unobservable event of interest, and then acting on the basis of that posterior.

For the problem at hand, the brain's initial uncertainty about the territory it must navigate can be expressed by a prior distribution over that territory $p(X)$, and the map it uses to guide it can be associated with a conditional distribution $p(Y|X)$. Together these yield a joint distribution $p(X, Y)$ which can be used to calculate informational quantities. It's a much bigger question whether (and how) these probabilities are explicitly represented in the brain. Cautious Bayesian cognitive scientists don't take a stance on exactly how probabilities are represented, rather asserting that the Bayesian calculus captures something important about how the brain deals with uncertainty (Perfors et al., 2011, pp. 302–303). A strong realist view would assert that all these distributions are explicitly represented and computed over. A more modest instrumentalist perspective suggests that prior uncertainty can be attributed on the basis of behaviour in the absence of evidence. For example, if a creature has no cognitive map and no sensory guide, how does it search the territory for a particular goal? If it searches randomly, its prior distribution can be represented as maximally uncertain.[11] Similarly its search behaviour upon being provided with evidence yields for the human experimenter a way to determine what likelihood should be associated with that piece of evidence.

My contention, in short, is that insofar as the cautious Bayesian approach is valid, both priors and likelihoods can be defined in principle in the case of cognitive maps. Priors and

[11] Indeed this is an assumption made by Haldane and Spurway (1954) and Wilson (1962) in their calculations of the informational quantities carried by the honeybee waggle dance and ant chemical trails respectively about the location of food.

likelihoods together define the joint distribution required to define informational measures like mutual information. It's a much larger question whether the Bayesian approach is indeed valid, and how it ought to be pursued (Jones & Love, 2011).

3.5 Example 5: binary strings

A surprising consequence of what I've said so far is that codes in standard communication theory are structural representations. The spatiotemporal ordering of binary strings in a transmitted signal corresponds to the spatiotemporal ordering of symbols in the source message that the signal represents.[12] If what I said in the taller-than case is correct, the orderings of digits in at least some binary strings are carrying information over and above the information carried by the digits themselves. And this seems like an undesirable consequence because we don't usually talk of the relations between digits carrying information. We say each digit carries 1 bit of information. That's why they are called bits: it's short for 'binary digit'.

Regardless how communication theorists tell the story, we are forced to follow the same logic as earlier. Knowing the relation between the digits of a binary string makes the difference between knowing only that there are e.g. two 1s and two 0s, and knowing whether it's 1100, 1010, 0011, etc. There are at least two reasons why we don't usually impute structural properties to binary strings. First, the relational structure is implicit in the spatiotemporal structure of the symbolic code. In the standard communication-theoretic scenario we already know exactly which digit belongs in which slot because we are receiving digits in the order in which they were sent. We don't come across evidence like 'there are two 1s and two 0s' without a specification of where in the code they stand. Even in cases of noise, that's not the form in which reconstructed evidence appears.[13] The fact that we aren't typically forced to consider the role of structural relations in binary strings is one reason they haven't been apparent. The second reason is that binary strings can be structured without being structural representations. Signs can have structure (comprising relations between their components) without that structure representing the structure of a signified. A binary string whose digit order does not correspond to an ordering in the referent of the string is not necessarily a structural representation. In the standard communication-theoretic model, the structure of a signal involves the ordering of binary strings within it, and this ordering does correspond to the ordering of symbols in the source message. Spatiotemporal ordering is assumed in typical models of communication theory.

3.6 Structure or information?

Part of what I'm trying to do is motivate the use of communication theory to understand representation. It seems as though there is an idea floating around that informational measures can deal with indicator states but not structural representations. I hope that I've at least started to put a dent in that idea. The assumption behind this idea appears to be that mutual information can only measure the strength of correlation between two processes that co-occur. Typical examples of indicator states carry information by co-occurring with their

[12] For an argument that communication-theoretic signals represent source messages, see Mann (2023).

[13] More complex real-world scenarios do have to confront the problem of out-of-order data. The transmission control protocol which forms part of the backbone of the internet is explicitly designed to account for and correct inadvertent reordering of data packets during transmission, along with many other functions improving connectivity.

signifieds, and mutual information is used to measure the strength of correlation between them. But the assumption that co-occurrence is required is false: mutual information applies to any two variables for which a joint distribution can be defined. A joint distribution could be defined for an indicator state that did not co-occur with its signified, but signalled its presence at a different time and/or place.[14] And a joint distribution can be defined for structured signs and their signifieds, because those probabilities need not be defined in terms of co-occurrence. The fact that a particular sign changes the probability of a particular signified need have nothing to do with the time at which the sign and signified occur. It is just that models of sign-reading behaviour typically do involve co-occurrence. Cases from psychology and cognitive science (e.g. Rescorla, 1988) and the models introduced by Skyrms (2010) are all like this. But it would be wrong to conclude, from the fact that models invoking mutual information usually involve co-occurrence, that models invoking mutual information must involve co-occurrence. Probabilities are not constrained like that. My map of India might change the probabilities of spatial relations between Indian cities as they are now, or if it is an old map it might change the probabilities concerning how these relations were in the past. The joint probabilities between signs and signifieds help us quantify how much can be learned about one from observing the other, rather than how often they co-occur.

Even supposing my reasoning so far is broadly correct, it is tempting to think that there is a better mathematical tool to employ when discussing structural representations. A measure called **Kolmogorov complexity** is often invoked in discussions of the information content of individual mathematical objects (Li & Vitányi, 2008, p. 2). Since we are modelling structured signs like cognitive maps as mathematical objects, and since these don't fall neatly into the Shannonian paradigm (see section 3.4), it might seem appropriate to use Kolmogorov complexity to model their information content.

Kolmogorov complexity captures how 'compressible' an object is. Slightly more formally, it is the length of the shortest computer program required to produce the object as output. The following example will clarify the concept. Consider two binary strings:

01010101010101010101010101
10111100100000101110111111

The first requires a relatively short program to produce, something like 'print "01" * 12' which prints the two-digit string "01" twelve times. By contrast the second string is much harder to compress. A computer program generating it would not be much shorter than the string itself. The first string therefore has lower Kolmogorov complexity than the second. By measuring the length of the shortest computer program required to produce a mathematical object, Kolmogorov complexity captures the compressibility or simplicity of objects. Since we have been modelling maps as mathematical objects, the measure applies to them too. The map of a completely bare territory requires a very short program to produce: the program instructions simply say to leave every square blank and output the resulting map. By contrast a richly flourishing territory with many different kinds of feature placed seemingly randomly requires a much longer program. The program must specify each feature's type and location. If the features are distributed randomly there will be no saving available from compression.

[14] Thanks are due to an anonymous reviewer for reminding me that indicator states need not co-occur with their signifieds, even though canonical examples do.

One might think that we should be using Kolmogorov complexity rather than mutual information to capture the informational content of structured signs. A rule of thumb says that static objects are measured with Kolmogorov complexity, while relations between random variables are measured with mutual information. Notwithstanding my earlier argument about Bayesian cognitive science, it might seem more appropriate to use Kolmogorov complexity rather than mutual information for representations like cognitive maps.

However, the rule of thumb that equates Kolmogorov complexity exclusively with individual objects cannot be taken too literally, because Kolmorogov complexity and mutual information are not mutually exclusive. Binary strings have a Kolmogorov complexity when they are treated as individual mathematical objects, but they also bear pointwise mutual information to source outcomes when they are employed as signals. So both types of measure apply, and they are capturing different things: one captures internal complexity of a sign, while the other captures relations between the sign and its signified.[15]

So the question we should ask is not 'Do structured signs have Kolmogorov complexity or mutual information?', but 'What do these measures capture, and how are they relevant to what we care about?' I have argued that mutual information captures the probabilistic relationship between the relations borne between components of structured signs and the relations borne between components of their signifieds. Insofar as we are interested in the mutual information between indicator states and signifieds, we should be interested in the mutual information between structured signs and their signifieds too. It's a further question what Kolmogorov complexity contributes.

Despite everything I've said so far in praise of mutual information, there are restrictions on its role and that of related measures in theorising about representation. Since the benefits of structural representations are often said to be format-specific, and the same amount of information can be carried by representations of different formats, informational measures alone cannot tell us what is beneficial about one format over another. The thought here is that there must be something beyond mere information-carrying that enables structural representations to offer more efficient or flexible processing to cognitive systems. In the next section I will flesh out this thought and assess the use of a complementary formal framework – computational complexity theory – to support it. I will conclude that informal arguments about the benefits of structural representation are not yet supported by the formal results of computational complexity theory. Whether another formal framework can cash out the benefits of structural representation I leave to future work. And whether we need a formal framework to do this job at all is something I cannot argue for at length here. At the least, I believe it would be worrying if there were no way to formally capture the benefits that structural representations intuitively afford. Others may disagree, but it is this belief that motivates my investigation for the remainder of the paper.

[15] Cover and Thomas (2006, §14.3) and Li and Vitányi (2008, §2.8.1) describe close formal relationships between Kolmogorov complexity and the measures defined by Shannon, especially entropy. I will leave to future work the question whether those relationships impact the claims I am making about mutual information.

4. Structural representation and computational complexity

4.1 Why structural representation?

The question that will occupy this section is: what are the benefits of structural representation? It looks as though any structured system of representation could in principle be replaced by an unstructured system that is carrying the same amount of information about the same signifieds. To construct such a system, arrange the structural representations in a list and replace each in turn by a unique unstructured representation. Keeping all the probabilities the same, we will end up with a set of unstructured representations that all bear the same information to their signifieds as did their structured counterparts. If every structured system is equivalent to an unstructured system in terms of information-carrying, why might cognition involve structural representations at all?

Informal answers to this question point toward the relative ease of carrying out certain representational operations in certain formats. As Coelho Mollo & Vernazzani suggest: "More appropriate formats will typically involve fewer, less complex, and less expensive computations than less appropriate ones" (Coelho Mollo & Vernazzani, 2023, pp. 16–7). It does seem intuitive that certain computations are easier to carry out with structured signs. Because maps represent spatial relations with spatial relations, adding a new location to a map automatically induces new relations between the new point and all existing points. If the relational information in a map were instead stored as a list of relations between points, adding a single new point would seem to necessitate adding many new entries to the list, each describing the relation of the new point to one of the existing points.

However, we ought not rely solely on our intuitions to evaluate arguments about the benefits of different representational formats. Pylyshyn (2002), for example, has demonstrated the difficulty of establishing the claim that mental imagery is a special format employed for image-based thinking.[16] Although our task is easier – stating the benefits of a certain format rather than establishing it is actually in use – the warning is nonetheless salient. One way in which we might want to rigorise the intuitive idea is by appeal to computational complexity. This section will therefore consider the possibility of a computational complexity analysis of structural representation for cognitive science. I will argue that the standard application of computational complexity to cognitive science, in which it is used to constrain computational-level hypotheses, is not suitable for establishing the benefits of structural representation. I instead suggest applying computational complexity at the algorithmic level of description.

Before getting started, I accept that the topic of this section may strike the reader as out of step with the rest of the paper. I want to insist that there is at least a common ideology between the information-theoretic perspective in the foregoing and the computational complexity theory perspective below. On the one hand, philosophers have previously assumed (or at least implied) that informational measures apply only to detectors or indicators, and that a different kind of relationship applies to structural representations; I have argued that a careful application of information theory reveals that view to be unsupported. Similarly,

[16] A reviewer pointed out an argument from Dennett (1991, §4.7), who describes a situation in which an artificial system is able to complete tasks that intuitively require image processing without employing internal iconic representations at all (where iconic representation falls broadly within the category of what I am calling structural representation).

we are about to see that philosophers assume that different representational formats enable easier performance of computational (including cognitive) tasks; I will argue that careful application of computational complexity theory reveals that that claim is similarly unsupported by formal work. The common method here is the application of formal theory to assess informal assumptions about representation. In the two parts of the paper we are investigating the same kind of representation – structural representation – through the lens of two different formal theories, information theory and computational complexity theory. By the end of this section it will be clear that the approach raises important questions about the benefits of structural representation. Overall, formal methods can complement and illuminate non-formal approaches in the philosophy of cognitive science.

4.2 Computational complexity

To understand computational complexity we must first understand computational problems. A **computational problem** is a function $f : I \to O$. A function can be thought of as a problem because for each input $i \in I$, one can ask the question 'what is the correct output $o \in O$ such that $f(i) = o$?' The development of algorithms that can solve computational problems – that is, that can instantiate computational functions – is an important activity in computer science. And the classification of problems according to their difficulty is an important theoretical pursuit, where 'difficulty' can be expressed in the precise formal terms of computational complexity.

Computational complexity is a property of a computational problem. Roughly, it captures the amount of computational resources required to solve the problem. Examples of computational resources are time – not real time measured in seconds, but a computational notion of time measured in terms of the number of computational steps required – and space – not physical space measured in metres, but a computational notion of space measured in terms of the amount of memory required. In order to precisely measure computational steps and memory, a specific model of computation must be employed. Typically the Turing model of computation is used, on which one operational step of a Turing machine counts as one time step, and each tape cell written to counts as one memory unit. More specifically, measures of computational complexity are expressed in terms of how resource usage grows as input size increases. This is reflected in the use of big O notation: the big O describes, for a problem, which function of its input size provides an upper bound on the time taken to solve the problem.

Consider for example the problem of multiplying a positive integer by 2. The input and output domain are the same, namely the positive integers. Multiplying an input of size n by 2 is the same as adding two n-digit numbers together, and the time complexity of this problem is known to be $O(n)$. This means that the number of steps an algorithm must take is at most cn where c is a constant. Although this might sound like a lot – both n and c can in principle be anything – it is generally the case that problems of this complexity can be feasibly computed for the kinds of inputs we're interested in.

The informal notion of feasibility just mentioned is a key concept in computational complexity theory and is more commonly known as tractability. A **tractable** problem is, informally, one that can be solved using a reasonable amount of resources. Problems are considered tractable when their complexity is a polynomial function of the input size. Problems whose complexity is larger than this – say, an exponential function of the input size – are

considered **intractable** (van Rooij, 2008, p. 947). Surrounding the divide between tractable and intractable problems are a variety of **complexity classes**, each characterised by a big-O function associated with the problems in that class. The complexity classes form a hierarchy: as one moves up the hierarchy, more and more problems are solvable in the time specified by the big-O function at each level (see Aaronson, 2005, for a compendium of complexity classes and known results).

The complexity of a problem may differ from the complexity of an algorithm used to solve it. A problem's complexity for a given resource is equal to that of the most efficient algorithm that solves it (and there may be no single algorithm that simultaneously optimises time and space). While a problem's complexity for a given resource can never be greater than that of an algorithm that solves it, an algorithm's complexity may be much higher than that of the problem it solves. The algorithm might solve the problem in a needlessly inefficient way. I will call an algorithm **inefficient** if it takes more resources than necessary to solve a problem. While some algorithms might be intractable, the tractable ones may be more or less efficient. This terminology captures the fact that different algorithms can use different amounts of resources to solve the same problem.[17]

4.3 Computational complexity in cognitive science

The past few decades have seen growing interest in applying computational complexity theory to cognitive science. The basic idea is to evaluate hypotheses at the computational level of Marr's hierarchy (Marr, 1982, p. 25) by considering the computational complexity of the function that the hypothesis attributes to the brain (van Rooij, 2008). If a computational-level hypothesis attributes to the brain the ability to perform a function that turns out, when appropriately formalised, to be intractable, then we can be confident the brain does not perform that function (at least, not for indefinitely many inputs of increasing size). This method of constraining computational-level hypotheses has been defended by Frixione (2001), van Rooij (2008), and van Rooij et al. (2019). Although this work is geared towards constraining hypotheses, there are other potential applications. van Rooij (2008, p. 967) describes testing algorithmic-level hypotheses by comparing reaction times of subjects with time complexity of the proposed algorithm. In that case modified versions of the complexity measure might be more appropriate: computational complexity takes the worst-case scenario for the amount of resources consumed in solving a problem for all possible inputs, but comparisons with human subject performance might warrant measures of the average case instead (van Rooij, 2008, p. 967). Whichever measure is used, measuring the complexity of algorithms offers a more fine-grained way to determine what the brain is doing when it solves a problem.

Can these considerations help answer the structural representation question? First the bad news. The popular approach that seeks to constrain computational-level hypotheses by rejecting intractable problems is not going to help answer our question. That's because the kinds of problems that are being solved by structural representation in cognitive science are thought to be equally computationally complex no matter what format you use (van Rooij,

[17] There are several interesting connections between computational complexity, which is a property of problems or algorithms, and Kolmogorov complexity, which is a property of mathematical objects (Li & Vitányi, 2008, §1.7.4). Confusingly, the term 'algorithmic complexity' has been used for both Kolmogorov complexity and the computational complexity of algorithms. Unfortunately I don't have space to go into further detail about these connections. As will become clear, there is a great deal more work to be done on the issues raised in this section, including on the relevance of the connection between the two concepts for the question of structural representation.

2008, p. 945). If this point extends to all possible formats – and that is generally considered a likely outcome, known as the Invariance Thesis (Frixione, 2001, p. 386), (van Rooij, 2008, p. 945) – then a problem that is intractable when coded non-structurally will not become tractable when coded structurally. The notion of structural representation occurs in computer science under the heading of analog computation,[18] and there appear to be inherent trade-offs when switching between analog and digital formats. Savings enjoyed with respect to one resource entail expenditure on another. For example, Vergis et al. (1986) describe a machine that compares the size of two positive integers using particles of two different masses. The authors show that for very large numbers that are very close together, the machine would have to be extremely large in order to enable discrimination between the masses. While a digital computer would require a lot of time to compare the sizes of two large integers, an analog machine could save time but only at the expense of space. There appears to be general agreement that trade-offs of this kind are the rule at the computational level.

One ray of hope might come from the topic of parameterized complexity. It turns out that certain hard problems can be cajoled into tractable form for restricted input domains. The definition of complexity looks to the worst-case scenario for all inputs of each size when determining how resource usage grows with input size, in effect treating all inputs of a given size as equivalent. The field of parameterized complexity instead distinguishes inputs based on the values of particular parameters they instantiate. In this way it is possible to determine a restricted set of inputs for which resource usage does not grow inordinately fast, even if usage grows too fast across all the inputs. The hope is to show that, for an hypothesised computational function, the actual inputs encountered by brains belong to this restricted class, enabling brains to employ a more tractable solution. Of course, if a brain encountered an input that was in the difficult class, it would be unable to solve it in a reasonable amount of time. But if the inputs it typically encounters are nice, then we can safely posit an otherwise intractable function after all. van Rooij (2008) and van Rooij et al. (2019) describe several seemingly intractable problems whose inputs can be parameterised in this way, thus widening the field of potential computational-level hypotheses.

While parameterized complexity is good news for cognitive scientists offering computational-level hypotheses, it will not help us answer our question. We are not positing intractable decision problems for which we want to find a restricted class of nice inputs. Instead, we are positing that one representational format is in some sense more efficient than another for certain sets of problems. If there was a problem which non-structural representation rendered intractable, but structural representation rendered tractable for parameterised inputs, then there would potentially be an application for parameterised complexity here. But just as computational complexity itself seems to be equivalent across formats – at best, savings on one resource lead to expenditures in another – the savings available from parameterisation appear to be equivalent too. van Rooij (2008) mentions parallel computers and quantum computers as obeying the same results, and to my knowledge there has been no suggestion that analog machines are exceptions to the rule. The conclusion seems to be that since complexity class membership isn't format-relative, answering questions about why certain formats are beneficial can't appeal to complexity class membership. In particular, it can't appeal to tractable-vs-intractable class membership.

[18] See for example Maley (2023), who offers an account of analog computation in terms of analog representation and suggests it generalises to an account of structural representation, and O'Brien and Opie (2015), who treat structural and analog representation as synonymous.

Now for the good news. Focusing on algorithms instead of computations opens up a more fine-grained set of questions about representation. If computational complexity is going to help, I suggest it will be at the algorithmic level rather than computational. A problem for which there are only inefficient algorithms employing unstructured representations might have efficient algorithms employing structural representations. In order to spell this out, we would need to regiment and formalise algorithmic-level hypotheses in the same way van Rooij et al. (2019, Appendix C.2) regiment computational-level hypotheses. It will be necessary to use a model of computation that supports describing algorithms using both structural and unstructured representations, or at least to use models that can be compared, in the same way that Vergis et al. (1986) (mentioned above) compare analog and digital machines solving an integer-comparison task. Only when algorithms can be fairly compared in terms of efficiency will we be confident that our claims about the benefits of structural representation are sound.[19]

It might turn out that some combination of algorithmic efficiency and parameterized complexity is key to answering our question. Perhaps algorithms employing structured representations are more efficient because they more faithfully delineate the nice inputs from the difficult ones. This tempers my argument above that parameterized complexity does not help: although it cannot help at the computational level, it might help at the algorithmic level. In general we should draw on all the formal resources we can in order to answer our question and justify or refute our intuitions.[20]

4.4 Summary and future directions

Mainstream applications of computational complexity in cognitive science aim to constrain plausible computational-level hypotheses. That isn't our task. It is generally accepted that a computational problem belongs to a particular complexity class regardless of how it is encoded. Therefore different formats cannot place a problem into different complexity classes. For these reasons, the mainstream approach to complexity in cognitive science will not help answer our question about the benefits of structural representation. Rather, if there is an answer to the question why structural representation is used, it will concern the complexity of algorithms rather than computational problems.

To apply computational complexity (whether at the algorithmic or computational level) we need to model cognitive tasks as functions $f : I \to O$. We need to explicitly specify inputs I and outputs O, and we need to be able to measure the size of an input $i \in I$. If there are tasks that cannot be modelled neatly in this way, we need to develop a wider notion of complexity. It is not yet clear whether the claim of Coelho Mollo and Vernazzani (2023, pp. 16–7), that different formats enable "fewer, less complex, and less expensive computa-

[19] Maley (2024) argues that the algorithmic level of Marr's hierarchy does not apply to analog representation (which here includes structural representation; see footnote 18 on page 22). On his view, whereas digital computations are carried out by algorithms, analog computations are carried out by mechanisms. If this is right, in order to achieve the task I am advocating in the main text, we must first figure out how to compare the efficiency of algorithms to the efficiency of mechanisms. This raises the question whether they are commensurate at all. The Invariance Thesis suggests that they must be, somehow; beyond the trade-off described by Vergis et al. (1986), I don't know of work that says exactly how.

[20] A reviewer pointed out that there may be systems that entirely lack the ability to process certain representational formats. In these cases computational complexity would not fully explain why the system uses the formats it does. We would have to appeal to the further fact that certain formats cannot be used by the system at all.

tions", is backed up by formal theory.[21]

One aspect not yet mentioned is the error-tolerance of structural representations. A small error in a structural representation can lead to a small error in output. Since computational complexity typically treats computations as yielding strictly correct or incorrect answers to problems, it seems to lack the formal tools to describe this feature. Although there are concepts of bounded error (Aaronson, 2005) and discrimination expenditure (Vergis et al., 1986), whether and how these correspond to the notion of error-tolerance as informally understood for cognitive representations is a question for the future.

Furthermore, once we're looking at cognitive systems, we see a third important class of resources after time and space: metabolic expenditure. It is not obvious that the time steps delineated in the Turing model correspond one-to-one to the metabolic costs of a brain carrying out those steps. A model of computation might inadvertently attribute identical costs to two different algorithmic steps that in fact consume wildly different amounts of metabolic resources when instantiated in the brain. Therefore, to gain an accurate picture of the benefits of structural representation, our computational model ought to attribute realistic resource costs to the algorithmic (or mechanistic; see footnote 19) steps it encompasses.

In general, there is a lot more than just information-carrying to consider when applying formal concepts to representation in cognitive science.

5. Conclusion

Structured signs carry information about their signifieds just as unstructured signs do. I have used simple examples to demonstrate how the information carried by a structured sign can be decomposed into the information carried by its components and the relations between them. The account relies on a relativist notion of probability; it is an open question whether a similar story could be told from an objectivist perspective. Finally, informal, plausible claims about the efficiency enabled by structural representation have not yet been cashed out in formal terms. I believe we ought to try to formalise the algorithms attributed to brains in order to explicitly measure the improved performance structural representations intuitively enable.

Acknowledgements

Thanks to Manolo Martínez for invaluable discussion on all aspects of this paper, especially forcing me to think harder about the various forms of complexity. Comments from two anonymous reviewers also greatly improved the manuscript.

This work was supported by Juan de la Cierva grant FJC2020-044240-I and María de Maeztu grant CEX2021-001169-M funded by MICIU/AEI/10.13039/501100011033.

[21] A reviewer pointed out that Coelho Mollo and Vernazzani (2023, p. 16) refer specifically to "[m]ore appropriate formats"; that is, formats that are more appropriate to solving a specific cognitive task. Given the authors' characterisation of 'more appropriate' in terms of consuming fewer resources, it seems that their contention that more appropriate formats will enable "fewer, less complex, and less expensive computations" is true by definition. That's as may be: my complaint is that their contention that different formats can *be* more or less appropriate has not yet been endorsed by formal theory. I entirely agree with the intuition, I simply think we should draw on formal tools to cash it out.

References

Aaronson, S. (2005). *The Complexity Zoo*. Retrieved January 25, 2024, from https://cse.unl.edu/~cbourke/latex/ComplexityZoo.pdf

Birch, J. (2014). Propositional content in signalling systems. *Philosophical Studies*, *171*(3), 493–512.

Burge, T. (2010). *Origins of Objectivity*. Oxford University Press.

Coelho Mollo, D., & Vernazzani, A. (2023). The Formats of Cognitive Representation: A Computational Account. *Philosophy of Science*, 1–20.

Cover, T. M., & Thomas, J. A. (2006). *Elements of Information Theory* (2nd ed.). John Wiley & Sons.

Dennett, D. C. (1991). *Consciousness Explained*. Little, Brown and Co.

Dretske, F. I. (1981). *Knowledge and the flow of information*. MIT Press.

Dretske, F. I. (1988). *Explaining Behavior: Reasons in a World of Causes*. MIT Press.

Favela, L. H., & Machery, E. (2023). Investigating the concept of representation in the neural and psychological sciences. *Frontiers in Psychology*, *14*(1165622).

Frixione, M. (2001). Tractable Competence. *Minds and Machines*, *11*(3), 379–397.

Gallistel, C. R. (2020). Where meanings arise and how: Building on Shannon's foundations. *Mind & Language*, *35*(3), 390–401.

Gutknecht, A. J., Wibral, M., & Makkeh, A. (2021). Bits and pieces: Understanding information decomposition from part-whole relationships and formal logic. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *477*(2251), 20210110.

Hájek, A. (2007). The reference class problem is your problem too. *Synthese*, *156*(3), 563–585.

Haldane, J. B. S., & Spurway, H. (1954). A statistical analysis of communication in "Apis mellifera" and a comparison with communication in other animals. *Insectes Sociaux*, *1*(3), 247–283.

Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge University Press.

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research*, *49*(10), 1295–1306.

Jones, M., & Love, B. C. (2011). Bayesian Fundamentalism or Enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, *34*(4), 169–188.

Li, M., & Vitányi, P. (2008). *An Introduction to Kolmogorov Complexity and Its Applications*. Springer.

MacBride, F. (2020). Relations. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2020). Metaphysics Research Lab, Stanford University. Retrieved January 17, 2024, from https://plato.stanford.edu/archives/win2020/entries/relations/

Maley, C. J. (2023). Analogue Computation and Representation. *The British Journal for the Philosophy of Science*, *74*(3), 739–769.

Maley, C. J. (2024). Computation for cognitive science: Analog versus digital. *WIREs Cognitive Science*, e1679.

Mann, S. F. (2023). The relevance of communication theory for theories of representation. *Philosophy and the Mind Sciences*, *4*.

Marr, D. (1982). *Vision: A Computational Investigation Into the Human Representation and Processing of Visual Information*. MIT Press.
Mehta, P., Goyal, S., Long, T., Bassler, B. L., & Wingreen, N. S. (2009). Information processing and signal integration in bacterial quorum sensing. *Molecular Systems Biology*, *5*, 325.
Millikan, R. G. (2013). Natural information, intentional signs and animal communication. In U. E. Stegmann (Ed.), *Animal Communication Theory* (pp. 133–146). Cambridge University Press.
O'Brien, G., & Opie, J. (2015). Intentionality Lite or Analog Content? *Philosophia*, *43*(3), 723–729.
Pandey, B., & Sarkar, S. (2017). How much a galaxy knows about its large-scale environment?: An information theoretic perspective. *Monthly Notices of the Royal Astronomical Society: Letters*, *467*(1), L6–L10.
Perfors, A., Tenenbaum, J. B., Griffiths, T. L., & Xu, F. (2011). A tutorial introduction to Bayesian models of cognitive development. *Cognition*, *120*(3), 302–321.
Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences*, *25*(2), 157–182.
Ramsey, W. M. (2007). *Representation Reconsidered*. Cambridge University Press.
Rathkopf, C. (2017). Neural Information and the Problem of Objectivity. *Biology & Philosophy*, *32*(3), 321–336.
Rescorla, R. A. (1988). Pavlovian conditioning: It's not what you think it is. *The American Psychologist*, *43*(3), 151–160.
Richmond, A. (2023). Commentary: Investigating the concept of representation in the neural and psychological sciences. *Frontiers in Psychology*, *14*.
Scarantino, A. (2015). Information as a Probabilistic Difference Maker. *Australasian Journal of Philosophy*, *93*(3), 419–443.
Shannon, C. E. (1948). A Mathematical Theory of Communication (Part 1). *Bell System Technical Journal*, *27*(3), 379–423.
Shea, N. (2018). *Representation in Cognitive Science*. Oxford University Press.
Skyrms, B. (2010). *Signals: Evolution, Learning, and Information*. Oxford University Press.
van Rooij, I. (2008). The Tractable Cognition Thesis. *Cognitive Science*, *32*(6), 939–984.
van Rooij, I., Blokpoel, M., Kwisthout, J., & Wareham, T. (2019, April 25). *Cognition and Intractability: A Guide to Classical and Parameterized Complexity Analysis*. Cambridge University Press.
Vergis, A., Steiglitz, K., & Dickinson, B. (1986). The complexity of analog computation. *Mathematics and computers in simulation*, *28*(2), 91–113.
Wilson, E. O. (1962). Chemical communication among workers of the fire ant Solenopsis saevissima (Fr. Smith) 2. An information analysis of the odour trail. *Animal Behaviour*, *10*(1–2), 148–158.

Stephen Francis Mann is a postdoctoral researcher in the Department of Linguistic and Cultural Evolution at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany. His work in philosophy focuses on the application of mathematical tools to philosophical problems in biology and cognitive science.

Address: Department of Linguistic and Cultural Evolution, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, Leipzig 04103 Germany
E-mail: stephenfmann@gmail.com – ORCID: 0000-0002-4136-8595

THEORIA
An International Journal for Theory, History and Foundations of Science