Induction and Inherent Similarity

T. Mark Ellison
Centre for Cognitive Science
University of Edinburgh
(MARKE@COGSCI.ED.AC.UK)

Abstract

Our perceptions are often logically compatible with abstractions we would never imagine entertaining. The problem of induction is to account for this disparity: how does evidence confirm one generalisation to the exclusion of others with which it is also logically compatible? In particular, how do we justify uniformity, the claim that the future will be like the past?

This paper introduces the problem of induction, and then proposes a solution based on similarity measures and topographic mapping. The premises of this solution are the following.

(i) All variation occurs in a context of non-trivial similarity structures.
(ii) Natural cognitive mappings between spaces of representation are topographic mappings, maximally preserving the similarity structures. The uniformity presumption can be justified by these two premises.

In this paper, I explore a new account for the difference between logical compatibility and inductive confirmation. It has been known for a long time that no amount of past evidence can assure any conclusion about the future (see the section Dark matter). Nevertheless, we reliably adopt certain potential generalisations and ignore others with conflicting predictions. The recognition of this selection and its lack of basis in logic is well recognised in philosophy where it is termed the problem of induction.

The problem of induction is not only of concern for philosophers, however, but psychologists and cognitive scientists as well. From any scientific account of human inductive behaviour, reliable experimental results must be deducible from experimental conditions. Imagine an induction experiment in which the subject is presented with a large number of green circles and blue squares. Asked to state a generalisation covering seen and unseen data, the subject is likely to aver that all circles are green, and all squares blue. The problem of induction implies that scientific accounts of such an experiment must adduce knowledge beyond what is present in the data to account for the test subjects’ step from instances to generalisation.

Drawing on the parallel between the problem of induction and the disparity between the observed and calculated mass of the universe, I shall refer to this additional information as the dark matter of induction.

(E1) Data → Human Induction → Generalisations

Dark Matter

This paper offers a new account of the nature of the dark matter of induction.

The next section of the paper relates this work to the wider context of similarity and categorisation. The section Dark matter surveys centuries of philosophical discussion highlighting the inadequacy of deduction in coping with certain kinds of inference. A particularly important uninduced predicate, Goodman’s grue, is introduced in this section, and some of the suggestions for eliminating it are surveyed in the section Stabs into the dark.

The solution to the problem of induction proposed here relies on the notion of topographic mappings. This is introduced in the section Topographic mappings and applied to the problem of grue in the section Untopographic grue. The paper ends with a discussion of the psychological force of the solution, and the reliability of learnability arguments.

Context

Here are some comments on the how this work relates to the twin fields of similarity and categorisation.

This paper is equally about both categorisation and simplicity. It deals with two kinds of categorisation, referenced by the terms attribute and predicate. Attributes are inherent properties of objects, defining a priori categories in the space of objects. In contrast, predicates characterise categories constructed through the processes of induction.

This dual system of categories is necessary for the easy construction of the concepts of philosophical discussion: truth, falsehood, predicates, and entities identified by their attributes.

One function of attributes is to define a similarity/difference measure in the data space. Similar objects are those which differ in few attributes: an unexamined green emerald is more similar to an examined green emerald than it is to an examined blue emerald or to examined green grass. The mathematics of this measure are discussed in the section Probabilistic similarity.

The only means to distinguish objects is their attributes,
and so objects which do not differ in their values for any attributes are taken to be identical.

It is the contention of this paper that induction in natural contexts exploits the similarity structure of the data-space, reflected in the system of attributes. Conversely, any data-space equipped with such an inherent similarity measure, offers its own unique possibilities for induction. While these claims are very general, a specific example will be considered, and so the abstractness of the argumentation will vary.

**Dark matter**

It is not a new realisation within the field of epistemology that logical validity, verified by arguments to contradiction, is the litmus test for certain knowledge. But it has been known for at least as long that much of our useful knowledge cannot be logically validated; it relies on something extra, the *dark matter of induction*. This section shows some of the history of our awareness of this gap between utility and certainty.

**Epicurus and Autrecourt**

According to Asmis (1984:197ff) the Epicureans\(^2\) recognised two forms of inference. One, counter-witnessing is akin to checking logical necessity by falsifying the converse. The alternative form, the method of similarity, allowed the similarity of objects to be used as a basis for ascribing them similar untested properties. The Epicureans, therefore, felt the need to make inferences which could not be justified by what we now call logical compatibility.

Nicolaus of Autrecourt was a 14th century scholastic who sought to temper his contemporaries’ enthusiasm for Aristotle by arguing that none of the Aristotle’s views could be regarded as certain\(^3\). He claimed that the only certain knowledge was gained by argument to contradiction. In contrast, habit-formation allows us much useful but uncertain knowledge, including all knowledge about the future (see Weinberg 1948:69). For him, therefore, the dark matter of induction was habit formation.

**Hume’s problem**

Hume’s name is attached to the problem of induction because of his exacting analysis and precise statement of the difficulty.

It is impossible, therefore, that any arguments from experience can prove this resemblance of the past to the future; since all these arguments are founded on the supposition of that resemblance (Hume 1975:38).

In fact, Hume suggested that at least part of the dark matter of induction might be similarity. He stated that all arguments on the basis of experience must be founded on similarity, and our expectation that similar effects flow from similar causes (p36).

---

\(^2\)Epicurus (341-271 BC) founded a school of philosophy which was active for approximately five centuries after his death.

\(^3\)For his pains, Autrecourt was forced to recant and witness the burning of his books, and also deprived of his degrees.

---

**Hempel and Goodman**

In an article on confirmation, Hempel (1945) considers the problem constituted by the following four statements.

**Raven 1 (R1)** All ravens are black.

**Raven 2 (R2)** All non-black things are non-ravens.

**Nicod’s Rule (NR)** A formula of the form \(A \text{ entails } B\) is confirmed by all instances logically compatible with it\(^4\).

**Equivalence Condition (EQ)** Whatever confirms (disconfirms) one of two equivalent sentences also confirms (disconfirms) the other.

The problem lies in *red shoes*. According to NR, red shoes are a confirmatory case of R2. But R1 and R2 are equivalent, and so EQ then requires that red shoes be regarded as confirming that *all ravens are black*, a counter-intuitive conclusion.

Hempel supports both NR and EQ, and claims that in this instance intuition has failed: red shoes are in fact confirmed, however weak for R1. In attempting to save NR, he hopes to maintain logical compatibility as a basis for confirmatory power. Goodman destroyed that hope.

Goodman (1954) introduced a new riddle for induction, showing that some predicates cannot be confirmed even though they share the form of R1. The riddle has also served to immortalise one of Joyce’s (1939:23) neologisms: *grue*.

Suppose a number of emeralds are examined before a certain time \(t\) and found to be green\(^5\). This evidence can be interpreted at time \(t\) as confirmation of the claim that all emeralds are green.

Now consider a second predicate, *grue*.

**Definition 1** An object is *grue* if its colour was examined before time \(t\) and found to be green, or if its colour was first examined after time \(t\) and found to be blue.

All our data, being about emeralds first observed before time \(t\), will report emeralds that are green, and consequently, grue. By NR, each of these observations supports the two statements: *emeralds are green*, and *emeralds are grue*. So the two generalisations have equal evidential support.

The two generalisations, however, make very different predictions about the first emerald to be examined after time \(t\). The green hypothesis asserts that it will be green, the grue hypothesis asserts that it will be grue, or equivalently, blue.

But these two predictions are not equal acceptable: we happily agree with, and make predictions from, *green* but not

---

\(^4\)Hempel ascribes this statement to the work of Nicod (1930), hence the name I have given it.

\(^5\)We here eschew all mineralogical knowledge to the effect that emeralds are *defined as green beryls* (Thomson 1966), and assume that emeralds could be identified in some achromatic, perhaps chemical, manner.
grue. In technical terminology, we say the green is projectible, while grue is not.

The riddle of grue is this: on what basis do we consistently exclude grue and aver the predictions of the equally-supported green hypothesis? In other words, what is the dark matter of induction? To this question, there is, as yet, no final generally accepted answer (Stalker 1994:2).

In this section we have seen the extent of the recognition of the gap between logical certainty and intuitions on induction. The next section looks at some of the responses which have sought to solve the riddle of grue, characterise the dark matter of induction and thus fill the gap.

Stabs into the dark

The lack of consensus on the solution to grue is, as one might expect, not for want of alternatives. These can be grouped into three major categories according to whether they regard the dark matter of induction as an intrinsic property of the data.

Extrinsic Which predicates are projectible depends on (at least partially) extrinsic properties of the predicate or data.

Goodman’s (1954) own solution to the problem posed by grue relies on an intrinsic distinguishing property on predicates, entrenchment. The disparity in induction can be resolved by looking at the past history of the predicates: green has been projected many more times than grue, a property which we say makes it more entrenched. In any induction process, such as the examining of emeralds, we opt for the more entrenched compatible hypothesis, ceteribus paribus.

The entrenchment solution seems to be merely asserting, unsatisfyingly, the conclusion of uniformity it should be justifying. The uniformity of our colour predictions for emeralds is justified by the uniformity of the choice of (entrenched) predicates we use to describe them.

Don’t care The action of the dark matter of induction, not its source, is what is important.

Sober (1994) offers a Bayesian analysis of the grue riddle. He points out that Bayes’ famous theorem offers no clue in and of itself as to whether green or grue is better. It states merely that before the distinguishing data arrives, any a posteriori preference for green can arise if and only if there is an a priori preference for green. In other words, the dark matter of induction must define a ranking of these a priori probabilities.

Sober seems content to leave his analysis there, having shown that if grue and green are not distinguished by the data, they must be distinguished elsewhere.

Intrinsic Which predicates are projectible is a consequence of properties intrinsic to the spaces of predicates and data.

Quine (1970, reprinted as 1994) identifies the dark matter of induction, and the solution to the problem of grue, with natural kinds. Green, he argues, is a projectible predicate because green things, or at least green emeralds, form a natural kind in a way that grue objects do not. The natural kinds are part of a system of similarity which arises from our innate perceptual system: it is because we are not colour-blind that we regard two red flowers as more similar than a red and a blue one. Nevertheless, through the progress of understanding, the native similarity measures can be replaced by more sophisticated ones. Thus similarity and natural kinds are intrinsic, but not inescapable, properties of perception and the space of values in which they arise.

This account of grue is deficient, however. The equation of projectibility predicates with natural kinds is asserted rather than derived. The following section introduces a model which combines Bayes’ theorem and similarity measures to explain rather than assert the preference for the green class.

A new approach

The approach adopted in this paper shares much of Quine’s response to grue. For example, it assumes a similarity measure (assumption 1).

Assumption 1 All variation occurs within the context of a similarity measure.

This assumption is about variation, be it in the value of a physical variable, in the data encoded in some representational system, or in the firing patterns of neural networks. All variation presupposes a space of alternatives from which differing values may be drawn; variants only have their identity by virtue of being distinct from the alternatives not chosen.

Any space of alternatives implies a similarity measure, ie, some relative notion of how distinct two variants are. The weakest measure is that needed to define the mathematical structure of sets: in which identical elements are equally self-similar, and distinct pairs of elements are equally dissimilar.

In contrast, pairs of real numbers show great differences in similarity indicated by the distance between them. The greater the distance, the less similar the points, a correspondence reflected in the definitions of continuity and differentiability.

Whenever variation occurs within a set of values, with no further substructuring, the corresponding set similarity measure is implicitly presumed. Whenever real numbers are used, the distance measure on real numbers provides the similarity measure unless it is explicitly overriden. Whenever we encounter variation, we implicitly encounter the structure of the space in which the variation occurs.

A second component of this account relies on the notion of topographic map. A topographic map is a function between two spaces with similarity measures which to a significant degree preserves these measures: similar objects in one space are mapped to similar objects in the other. One function is said to be more topographic than another between the same spaces, if it is more consistent in matching the similarity measures in the two spaces. A formal account of topographic mapping will be given in the section Topographic mappings.
Topographic mappings play a vital role in the new account of induction\(^6\) (assumption 2).

Assumption 2 Given some data linking two representation spaces, the preferred generalisation of that data is the most topographic function between the spaces which is compatible with the data.

Predicates, like Emerald implies Grue and Emerald implies Green, are functions from a space of objects to a space of truth values. Given reasonable similarity measures on the spaces of objects and truth values, the requirement of topographicality suffices to prefer the second predicate. In this account then, similarity measures inherent in the data, together with the requirement for topographic mappings, are the dark matter of induction.

In the next two sections, this model is explored in more detail. First, the section Topographic mappings offers an exact Bayesian definition of what it means to be a topographic mapping. Second, the section Untopographic grue shows the action of this definition in excluding the predicate Emerald implies Grue.

### Topographic mappings

The intuitive notion of topographicality, namely the preservation of similarity, is quite accessible. Making precise arguments, however, requires a formal measure of this property.

Unfortunately, there is not single unarguable definition of a topographic mapping. The definitions that do exist appeal to the same intuitions, but except in certain pristine conditions (e.g. when the mapping is guaranteed to be bijective), this leaves room for considerable variation in mathematical form. Goodhill\&Sejnowski (1997) survey a number of definitions of topographicality and exemplify their effects on a dimensionality-reduction task.

### Probabilistic similarity

A firm basis for the definition of topographic mapping can be achieved by tying the notion of similarity to probability of replacement. We can say that two points in a space are similar if the probability of one being replaced by, or mistaken for, the other is high. For a space \( S \), we write the probability of replacing point \( y \) replacing point \( x \) as \( P_S(y \leftarrow x) \).

### Supervised topographic mappings

Supervised learning is about function acquisition. The learner has access to a collection of matching inputs and outputs. The acquisition task is to recover the function which generated these pairs. Put more formally, we denote by \( IO \) the bag (the training data may include repetitions) of known input-output pairs, and the potential completions of these pairs by functions \( F \). Bayes’ theorem then provides an evaluation for each \( F \) in terms of the known data (E2).

\[(E2) \quad P(F \mid IO) = \frac{P(IO \mid F) \cdot P(F)}{P(IO)}\]

If the space of functions \( F \) is finite, we can adopt the maximum entropy choice of \( a \) priori distribution, assigning each function equal probability. The generalisation which maximises (E2) will consequently be the one which maximises the derived probability of the data \( P(IO \mid F) \).

A reasonable assumption is that conditional probabilities of the given input-output items are independent of each other. The conditional probability of \( IO \) can thus be decomposed into the product of the conditional probabilities of the elements.

\[(E3) \quad P(IO \mid F) = \prod_{(i,o) \in IO} P((i,o) \mid F)\]

At this point the model of derivation is crucial. Without the effect of similarity, \( P((i,o) \mid F) \) would be zero if \( o \neq F(i) \) and one divided by the number of pairs accepted by the function, \( 1/|I| \), in all other cases. With similarity effects in both the input and output spaces, we must consider the effect of possible replacement of the data. For a pair \((i,o) \in IO\), there is a certain probability \( P_l(i \leftarrow j) \cdot P_o(o \leftarrow p) \) that it could be generated from another pair \((j,p)\) by replacement. We can ignore replacement from pairs not permitted by the function \( F \) as these are of zero direct conditional probability. Thus we should consider replacement from pairs of the form \((j,F(j))\).

Summing the probability of all of the ways of deriving a pair from the function, we have the conditional probability of the pair given the function (E4).

\[(E4) \quad P((i,o) \mid F) = \sum_{j \in I} \frac{P_l(i \leftarrow j) \cdot P_o(o \leftarrow F(j))}{|I|}\]

Summing the probabilities of the mutually exclusive ways in which \( o \) can be paired with \( i \) gives the total probability of the pair \((i,o)\) given \( F \). This leads to the expansion of \( P(IO \mid F) \) shown in (E5).

\[(E5) \quad P(IO \mid F) = \prod_{(i,o) \in IO} \frac{\sum_{j \in I} \frac{P_l(i \leftarrow j) \cdot P_o(o \leftarrow F(j))}{|I|}}{1}\]

The most topographic function \( F_{IO} \) on the basis of input-output pairs \( IO \) and the similarity functions \( P_l \) and \( P_o \) is the one which maximises \( P(IO \mid F) \) in (E5). In the following section, we will apply this formula to the evaluation of grue.

### Untopographic grue

We now know how to evaluate the topographic nature of functions on the basis of a collection of example forms. To apply this to the problem of grue, we must first identify the functions involved and the corresponding input and output spaces. Secondly, these spaces must be equipped with appropriate probabilistic similarity measures. Finally, the potential predicates, including green and grue must be evaluated to find the

---

\(^6\)Some of Hume’s statements presage this assumption. He (1975:36) describes as the basis of induction the belief that similar effects come from similar causes; in the present terminology, the presumption that cause-effect link to be a topographic mapping.

\(^7\)We assume here that replacement in input and output spaces are independent.
most topographic in terms of the known data. These steps are covered in the following three subsections respectively.

**Predicates as functions**

Predicates are functions from a space of objects to a binary set of truth values. For example, the predicate *is green*, when applied to the object *my emerald* returns the truth value true, meaning occurs, but when applied to the *Scottish flag* returns false, meaning doesn’t occur.

As described in the section **Context**, the space of objects is characterised by a system of attributes. These define the similarity structure on the space, and even identity — objects which share exactly the same set of attributes are identical. Given this role of attributes, there is little reason not to identify the space of objects with the product of the spaces of all possible attribute settings. If there are \( n \) attributes, and these can only be present or absent, then the space of objects forms an \( n \)-dimensional hypercube.

To use topographic mappings to model induction, we need to construct an object space rich enough for the predicates of the *grue* riddle. Thus the objects have to be distinguished by attributes *green/non-green, emerald/non-emerald, and examined/unexamined*. The predicate *Emerald implies Green* can therefore be defined as the function which takes all combinations of attribute values to the truth-value true, except non-green emeralds. In contrast, *Emerald implies Grue* takes all objects to true except examined non-green emeralds and unexamined green emeralds. The structure of the two predicates is shown in table 1.

<table>
<thead>
<tr>
<th>Emerald implies</th>
<th>K</th>
<th>nK</th>
<th>G</th>
<th>nG</th>
<th>E</th>
<th>nE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Green</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>E</td>
<td>nE</td>
</tr>
<tr>
<td>Grue IO</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

Table 1: The two predicates *Emerald implies Green* and *Emerald implies Grue*, together with the set of input data IO. Distinctive predicate values are shown boxed. \( K, \) \( nK, \) \( G, \) \( nG, \) \( E, \) and \( nE \) are abbreviations for the attribute values examined, unexamined, green, non-green, emerald and non-emerald. \( T, \) \( F \) and \( ? \) indicate true, false and unspecified respectively.

This account uses examined as an attribute of objects, in just the same way as greenness and emeraldhood are used. This is in keeping with the formulation of the problem in the philosophical literature.

**Similarity of objects and truth-values**

Now let us turn to the similarity functions on these two spaces.

There is a natural distance measure on objects spaces defined in terms of independent attributes. If an attribute \( a \) stands a chance \( P_E(a) \) of being errantly specified, then it has a chance of \( 1 - P_E(a) \) of being correctly specified. The probability that any given set of attributes will be correctly specified can be obtained by forming the product of the probability that each attribute will be correctly specified.

If objects are completely defined in terms of their attributes, this is sufficient to determine the probability of replacing one object with another. This is equal to the probability of not changing any of the attributes with a common specification for the two objects, multiplied by the probability of changing all the attributes in which they differ.

Let us formulate the above mathematically. We write \( A_{g12} \) for the set of attributes which are held by either both or neither of the objects \( o_1 \) and \( o_2 \), and \( D_{g12} \) for the set of attributes held by exactly one of the two. The probability \( P_E(o_2 \leftarrow o_1) \) of \( o_2 \) replacing \( o_1 \) is given by (E6).

\[
P_E(o_2 \leftarrow o_1) = \left( \prod_{a \in D_{g12}} P_E(a) \right) \left( \prod_{a \in A_{g12}} (1 - P_E(a)) \right)
\]

For the output space of a predicate, viz. the binary space of truth-values, the similarity measure is the same as for a single-attribute, characterised by one probability \( P_E \) of error and \( 1 - P_E \) for a single attribute distinguishing true from false.

**Simulation results**

The evaluation measure of the section **Supervised topographic mappings** was used to determine the most topographic predicates given the data on examined objects. The *a posteriori* probabilities of the ten best maps are shown in table 2.

<table>
<thead>
<tr>
<th>Predicate accepts</th>
<th>( P(F \mid IO) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Em \rightarrow Green: )</td>
<td>0.0289411</td>
</tr>
<tr>
<td>0 K G KG GE KGE</td>
<td>0.0269198</td>
</tr>
<tr>
<td>0 K G KG GE KGE</td>
<td>0.0266897</td>
</tr>
<tr>
<td>K G KG GE KGE</td>
<td>0.0264022</td>
</tr>
<tr>
<td>0 K G KG KGE</td>
<td>0.0248666</td>
</tr>
<tr>
<td>0 K KG GE KGE</td>
<td>0.0245661</td>
</tr>
<tr>
<td>0 K KG KG</td>
<td>0.0240565</td>
</tr>
<tr>
<td>0 K KG GE KGE</td>
<td>0.0243142</td>
</tr>
</tbody>
</table>

Table 2: The *a posteriori* probability of topographic predicates given the data IO shown in table 1. Predicates are specified by the attributes of objects which are mapped to the true truth value. Here \( 0 \) represents objects with no attributes set. Other labels as in table 1.

The most probable predicate is the one which accepts all combinations of attributes except non-green emeralds (E, KE), regardless of whether these have been examined or not. In other words, *Emerald implies Green* is the preferred generalisation. *Emerald implies Grue* is less favoured, being 15% less probable.
The margin by which *Emerald implies Green* dominates the alternatives depends on the number of independent exemplars in the already examined data. If 16 independent experiments show up only green emeralds, and non-emeralds of all colours, then *Emerald implies Green* grows ten times as likely as *Emerald implies Grue*. After 75 such experiments the *a posteriori* probability of *Emerald implies Green* increases to 99%.

This simulation has shown that *Emerald implies Green* is a more topographic mapping from objects to truth-values than *Emerald implies Grue*. Thus the stipulation that induction should extend data points to the most topographic compatible mapping automatically distinguishes *green* from *grue*, and favours the former.

**Discussion**

Let us retrace the argument in this paper. After introducing the current work and placing it within the discussion of similarity, a number of sources were adduced to highlight the gap between logical compatibility and inductive confirmation; a gap containing the dark matter of induction. Goodman’s riddle was seen to bring this disparity into particularly sharp focus.

Three responses to the riddle were surveyed. None of these proved completely successful in exposing the nature of the dark matter. An alternative was proposed based on two assumptions: that all variation occurs in spaces with implicit similarity structure, and that these spaces are best connected by maximally topographic mappings.

To assist the development of this account, a formal measure of the topographic nature of mappings was presented. This measure was used to show that *Emerald implies Grue* is an inferior generalisation to *Emerald implies Green*. This result indicates that intrinsic similarity measures and maximally topographic mappings are sufficient, without other dark matter, to account for human inductive behaviour.

**Psychological force**

An interesting question concerns the psychological force of this explanation of what makes a good prediction. *Grue* is a worse prediction than *green* is because it results in a lower probability for the data in terms of the predicate — see equation (E2). In other words, if you assume that the future is different from the past, then you will be less sure of the past. Thought experiments suggest that this is entirely reasonable.

Suppose that you have access to meteorological records for the town *Seca* dating back 250 years and these claim that it has never rained in that town. You decide to travel to see this marvel. When you arrive, you see rain in the near distance, and the clouds dropping it are moving rapidly towards the town. You find yourself with a strong belief that whatever its history, Seca, will see rain in the next 10 minutes. Does this affect the credibility of the records? Certainly. In accepting a prediction of rain, we are reducing the credibility of our past data.

For this reason, if we believe the records, then we will not predict a scenario like the one I have just outlined, because it would serve to diminish that belief. Instead we would expect a sunny, or at least, dry, environment in *Seca*.

So our predictions of the future affect the past, and thus accommodating that past means not only adapting our memory to fit, but our predictions as well.

**Why induction seems hard**

In the course of this paper, we have seen the effect of taking similarity information into account, in contrast to relying solely on logical compatibility. The green/grue dilemma was insoluble by compatibility alone, but presented no difficulty once predicates were required to be maximally topographic.

Parallel situations are likely to arise whenever the dark matter required for induction is disrupted or ignored. Consider the example of language. Language can be regarded as a mapping from meanings onto utterances. Children seem to acquire the mapping on the basis of exposure to words in contexts where their meanings can be deduced. And yet, like the bumble-bee’s flight, language acquisition seems impossible.

Gold (1967) idealised a model of language in which all utterances were either identical or absolutely distinct. A language consisted of no more than the set of allowable utterances, in effect the image of all possible meanings under the action of the language function. Gold showed that no learning strategy is guaranteed to be successful on learning all languages in even fairly modest classes of language, eg the class of all regular languages.

But let us look closely at what has happened to the model of language in the process of creating Gold’s model. All notion of degrees of similarity have been stripped from the space of utterances as well as the space of meanings. There is only identity, and absolute difference in each case.

As a consequence of this lack of similarity structure, no assumption of topographicality can be called on to help identify the language responsible for the data. The only criterion available, as was the case with Goodman’s version of green/grue riddle is logical compatibility. This criterion proves inadequate, and learners readily fail to identify the correct language.

In general, the disregard in formal models of the similarity structure available to natural learners results in over-strong claims about the difficulty of the task of induction.

**Why human induction is easier than it looks**

Formal models do not need, however, to disregard similarity structure. The green/grue case illustrates how we can introduce formal models which respect relatedness of values within a space of variation. Consequently, these permit distinctions between inductions which are not achievable through logical compatibility alone.

The condition that mappings be maximally topographic imposes an ordering on the examination of possible mappings. This means that more topographic mappings will be
tested earlier in the acquisition process. This makes in-duction easier so long as the target mapping is also topographic. If the target mapping is a human generalisation, then there is some reason for expecting to be topographic.

Firstly, there seems to be a neural basis to topographic mappings. Von der Malsburg (1981) developed a “correla-tion theory of brain function”. This theory proposes the funda-mental mechanisms of visual organisation are topographic mapping mechanisms, relating visual stimuli with structures stored in long-term memory.

At a psychological level, similarity and topographicality tie in with Lakoff and Johnson’s (1980) work on metaphor as an all-pervasive component of cognition. If similarity is defined in terms of directness of connection within schemas, topographic maps are metaphors linking schema so as to max-imise the correspondence of structure. The fundamental place accorded to topographic maps in von der Malsburg’s and sub-sequent work in neuroscience parallels the pervasive role for metaphor proposed by Lakoff and Johnson.

If topographic mapping is a fundamental human mecha-nism, as these studies suggest, then the ease with which chil-dren learn language may be explained. Children are exposed to input combining understood semantics with utterances, and the child apprehends each of these within the context of its own similarity space. As the child receives this input, it is stored, and topographically extended to cover the space of meanings the child wishes to express and understand: the full language mapping. When the child knows few words, the ef-fect of the topology will be small, as the data is spread so sparsely within it. As the number of words increases, topo-logical similarity effects will arise, and we can expect to see extrapolation, both correct and incorrect, in the language be-haviour of the child. As more data is acquired, these reinforce irregular structures, defeating over-regularisation.

The point here is that children are using topographicality, like their language informants, in the continuous process of learning and creating the language to describe new meanings. Because both rate possible languages according to the same criterion, the ability of the child to rapidly acquire the right language is unsurprising.

Could we mimic this kind of human learning in an artifi-cial computation system? The answer is yes, at least in part. Any learning method which makes use of richer similarity structures than sets is likely to offer improvement on current methods.

Humans learn from each other and from the world rapidly and easily because we attempt to match similar inputs to sim-ilar outputs. This is how other humans behave, and seems to be true of the world as well. Learning which ignores this heuristic is hard, and often goes unrewarded.

References


