Sets of Signals, Information Flow, and Folktales

Mark Alan Finlayson

Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 32 Vassar Street, Cambridge MA 02139, United States of America
markaf@mit.edu

Abstract. I apply Barwise and Seligman's theory of information flow to understand how sets of signals can carry information. More precisely I focus on the case where the information of interest is not present in any individual signal, but rather is carried by correlations between signals. This focus has the virtue of highlighting an oft-neglected process, viz., the different methods that apply categories to raw signals. Different methods result in different information, and the set of available methods provides a way of characterizing relative degrees of intensionality. I illustrate my points with the case of folktales and how they transmit cultural information. Certain sorts of cultural information, such as a grammar of hero stories, are not found in any individual tale but rather in a set of tales. Taken together, these considerations lead to some comments regarding the "information unit" of narratives and other complex signals.

1 A Theory of Information Flow

In their book "Information Flow: The Logic of Distributed Systems," Barwise and Seligman [1] present a mathematically sophisticated theory of *how things can carry information about other things*. Barwise and Seligman started from Dreske's seminal work on information flow [2], and expanded and formalized his observations, integrating his approach with related approaches, resulting in a more general formulation. (From here on out I will refer to this general formulation as the "DBS" theory of information flow, short for Dreske-Barwise-Seligman). I observe that the DBS theory is, in fact, even more general than it at first appears, and it is my aim to illustrate how it can be used to frame and describe several important facets of information flow, knowledge, and belief that were left unelaborated in both Barwise and Seligman's and Dreske's work. In particular, I will show how the DBS theory, without modification, can be used to conceptualize two important items which Dreske touched upon only tantalizingly: learning and intensionality. I show how this conceptualization brings into relief a part of information channels that is often taken for granted in philosophical analyses, namely, the process by which categories are applied to raw signals. I will then apply these insights to make some comments on the information content of cultural narratives (folktales).

I set the stage by reviewing in brief the relevant parts of the DBS theory. The theory involves, at its core, *classifications* and *infomorphisms*. These two

S.B. Cooper, A. Dawar, and B. Löwe (Eds.): CiE 2012, LNCS 7318, pp. 228–236, 2012.
© Springer-Verlag Berlin Heidelberg 2012

objects are used to model how information flows across *distributed systems*, which are systems that can be analyzed in terms of both a whole and constituent parts. In Barwise and Seligman's terminology, an *information channel* brings classifications and infomorphisms together into a full model of the information flow of a particular system.

We shall lose no generality if we restrict ourselves to discussing a distributed system W comprising only two parts, a proximal part P, to which we have direct access and be thought of as the "receiving" end, and a distal part D, from which information is flowing. There are infomorphisms that map properties of the classifications of the distal and proximal parts to the whole; call these d and p, respectively. To provide a concrete example to discuss, let us take Barwise and Seligman's example of a nuclear reactor: in this case W is the whole reactor, D will be the reactor core, and P will be a gauge in the reactor control room, and d and p are the regularities that connect the core to the reactor to the gauge.

A *classification* is similar to what one thinks of when considering the standard classification task in cognitive psychology: it is a set of labels or classes that may be applied to some object or phenomenon. Classifications can be, for example, mutually exclusive (e.g., {SQUARE,CIRCLE}), exhaustive (e.g., {TRUE,FALSE}), or overlapping (e.g., {TALL,FAT}). They can also be none of those things. Importantly, though, each part, as well as the whole, receives a classification. For our reactor example we might consider the reactor core D to be classified by the exclusive types NORMAL and OVERHEATING, the reactor status gauge can show one of GREEN or RED, while the reactor overall can be in one of the four states achieved by the cross product of these two classifications.

An *infomorphism* relates classifications on a part to classifications on the whole. It is a way of describing how classifications are transformed as the information they carry moves through the distributed system, from one part to another: they are models of the regularities that allow information flow. In such a system one infomorphism d may be applied to the distal part's classification to obtain a classification on the whole, and then another infomorphism p may be applied in reverse to the classification on the whole to obtain a classification on the proximal part. We need not say too much about infomorphisms except that, as they are applied in the forward and reverse directions, the resulting classifications loose some of their guarantees and internal relationships and are downgraded to what Barwise and Seligman call *local logics*. In the reactor example, the combined infomorphisms from reactor core to reactor whole, and then from reactor whole to control room gauge, given a reactor in working order, results in a display of GREEN on the gauge when the core is NORMAL, and a display of RED on the gauge when the core is OVERHEATING. Thus information flows from the distal part of the system (the reactor core) to the proximal part of the system (the control room gauge).

The details are not critical to my argument, but there are two essential points to take away from this description. First is that regularities across the system, modeled by chains of infomorphisms, are what allow information to flow from one part of the system to another. (It is often helpful to think of these regularities, like Dretske did, as lawful relationships, but one must remember that not all regularities are lawful.) Second, classifications are the language by which the information is communicated and information flow is relative to the classifications of the whole and its parts.

2 Information Flow via Sets of Signals

Information flow in the DBS theory is intimately connected, whether explicitly or implicitly, with *signals*. The examples in Dretske's and Barwise and Seligman's work are all concerned with individual signals. A signal is not defined precisely in either work, but one gathers it follows the natural intuitions: signals carry information and they are relatively localized in time and space. Signals flow across a distributed system from the distal part to the proximal part. They are the messages that contain the information. A light flashing SOS, reading the symbols off a map, a speech act: these are all signals.

I turn to an interesting and important case, that where the information of interest is not present in any individual signal, but rather is carried by correlations between signals. Regularities in a distributed system can result not only in information carried in a single signal; certain types of regularities can also result in correlations between signals.

How can this be so? Here is an example. Let us consider the reactor, and ask a simple question: Is the reactor more often NORMAL or OVERHEATING? Perhaps not a very interesting question, and one whose answer is obvious to anyone who knows much about nuclear reactors and how they are designed and run. But imagine that you know very little at all about nuclear reactors. Then, certainly, you would agree that if you were to learn the answer to this question about a particular reactor, then you would be the recipient of information. How we answer the question is straightforward: we simple observe the gauge periodically, noting whether it the gauge is NORMAL or OVERHEATING. Eventually, we stop and count up our observations, and whichever type outnumbers the other, that is our answer for this particular reactor.

How often we observe the gauge, for how long, doesn't matter much for my argument. What is important is that we cannot know, by observing any individual signal from gauge, whether the reactor is more often NORMAL or not. The information is not contained in any one signal, it is only contained in a collection of those signals. Now, one may object that this question is contrived and uninteresting, and does not represent the sort of information we are interested in studying. But, in fact, this is a common scientific question: "Is it more likely that X or Y for some type of signal?" Doctors, for example, ask the question of whether or not patients are more likely to die or be cured (or something in between) when they use or don't use a particular drug. Engineers ask whether a building is more likely to fall down in an earthquake when designed this way or that. Astronomers ask whether it is more likely for a type of star to go supernova sooner rather than later, or not all, if it has this or that characteristic. All of these examples are more complex than the reactor example, in that answering these questions usually involves correlating signals from multiple parts of the system at many different times, using much more complex methods, but the basic principle is the same: one cannot obtain the information from any one signal. The set of signals itself becomes an information channel.

How may this be analyzed within the DBS theory? The distributed system becomes, not a single instance at a particular time of the system under consideration, but a set of distributed systems, each one at different times. Each instance might be a specific time instance of a particular distributed system, they might be different instances of different distributed systems (all similar in some relevant way), or some mixture in between. The infomorphisms still reflect the regularities that underlie the system, they now just describe regularities spread across distributed system instances, and thus, time and space. The classification may be thought of as all the possible answers to the question — what those who do statistical analyses might call the *hypothesis space* of the problem. When we finally determine what is the actual answer, we have pinned a particular type on the receiving end of our set-of-signals distributed system, and we are the recipient of information.

This focus on the set of signals and the observation that scientists use sets of signals to answer scientific questions highlights an important fact: the way one correlates the signals in the set is key to the extraction of the information. Different methods result in different information flowing across the system. This choice of method contains much of the contribution of science: how do we process the raw data so as to uncover the information that we seek?

Naturally, the differences in information between methods may result from different classifications used or implied by each method. This is exactly the same as in the individual signal case where different information flows when we have a different classification for a part or the whole. But, there is an important distinction I would like to highlight, namely, that different methods might produce different answers for the *same classification*. For example, some correlation techniques might give a wider or narrower range for an answer (on an ordinal scale); on the other hand, a different technique might give a completely different answer. Thus in the scientific literature much effort is spent on justifying one's technique on principled grounds, and much is made of two different techniques converging on the same answer.

3 Learning and Intensionality

The above treatment shows that the DBS theory may be applied beyond examples containing a single signal. This allows us to frame two phenomena that are left unelaborated in the DBS analysis.

The first phenomenon is learning. Dretske noted that "Learning, the acquisition of concepts, is a process whereby we acquire the ability to extract ... information from the sensory representation." [3, p. 61] Learning can be described in the DBS theory by framing it as a set-of-signals information channel. We begin with individual signals that are unclassified. Moving up to the set of signals level, we apply a correlation method for identifying the type that applies to particular signals in particular circumstances. Having learned this classification we may return to the single signal case, and apply the newly learned classification. In the reactor example, suppose we learn, via observations at multiple times, and application of a particular correlation method, that certain gauges on the reactor always move in synchrony. This is a classification. When next presented with an individual signal, where perhaps we can observe only a single gauge, we can infer the state of the other gauges from a single observation. Similarly with what presumably happens when a child learns a new word. Daddy says "airplane!" and points. This happens several times. Perhaps there are some near-misses that aid learning ("No, honey, that's a butterfly.") Eventually, by correlations between all these signals, the child learns the category, and now can say "airplane" herself when seeing only a single signal. Learning has occurred.

The second phenomenon is degree of intensionality. Dretske said: "Our experience of the world is rich in information in a way that our consequent beliefs are not. ... The child's experience of the world is (I rashly conjecture) as rich and as variegated as that of the most knowledgable adult. What is lacking is a capacity to exploit these experiences in the generation of reliable beliefs (knowledge) about what the child sees. I, my daughter, and my dog can all see the daisy. I see it as a daisy. My daughter sees it simple as a flower. And who knows about my dog?" [3, p. 60] Dreske describes these differences in the perceptions as differences in *intensionality*. We can characterize this degree of intensionality by equating it with the method for extracting the information from the signal. The more sophisticated the correlation method, the more complex and varied the proximal classification, the more *intensionality* we assign to the agent in question. (This observation might lead us to hope that we can provide a full or partial order over intensionalities. Unfortunately this is not to be — see the next section.)

There are a number additional observations that may be made on this topic. For example, if we talk about information carried by sets of signals, why not talk about information carried by sets of sets of signals? Or sets of sets of sets? This, perhaps, is the same as talking about learning about learning, and so forth. We might also explore how the scientific method in general, or specific fields of inquiry, such as machine learning, are illuminated by these observations. We could examine in more detail how the learning method intervenes between signal and classification. But rather than explore these interesting lines of inquiry, I turn my attention to an application of these observations to a domain of particular interest to me: cultural information as carried by sets of folktales.

4 Folktales and Narrative Structure

My switchings gears to the topic of cultural information as carried by sets of folktales may seem like a *non sequitur*. I assert several reasons for the attention. First, narratives are an excellent example of a complex signal which contains myriad subtle sorts of information. Everyone is familiar with folktales, and so they will serve as an effective proxy for all sorts of complex signals with multiple possible interpretations without the overhead of detailed setting of the ground. I would like to use what I know about them to explore more this idea of varying degrees of intensionality, and for this purpose they have the advantage of us not yet knowing, scientifically speaking, what exactly is the information contained in them, and therefore we need not suspend our disbelief to imagine that there may be several ways of interpreting the information contained in folktales: we have several proposals (I will consider two) and we don't know which one (or ones) is right. Second, these observations allow me to pose, and explore a bit, some interesting questions about the nature of information carried in narratives. Third, narratives and culture are of central importance my work, and I am the one writing this paper. So bear with me.

Folktales specifically, and narratives in general, clearly communicate information aside from any considerations of their properties across a set. They are like any other text: they communicate information as individual objects. In a folktale in particular, and narratives in general, we can learn things such as who the characters are, what their plans and goals are, and what they are doing and when. (Although, usually being about a fictional world, it is an interesting question whether this information translates into knowledge.) This sort of information, the sort contained in an individual tale, is not the information we are interested in here. I am interested, rather, in information that is communicated across a set of tales.

There are numerous types of information communicated by a set of folktales. My work so far has focused on a particular sort, that of narrative structure of the plot [4]. This information corresponds to a grammar for plots, specific to the culture in question. Much like a natural language, a folktale grammar provides a set of symbols (plot pieces) and rules of combination that allow us to build folktales in that culture. Much like the grammar of a natural language is not captured in a single sentence, the grammar of the folktales is not captured in any single tale. There have been many proposals for the form of these grammars, proposals that span the range from universal theories across all stories, to highly culturally-specific theories for certain genres of folktales. I will contrast two examples, the first being Vladimir Propp's theory of the morphology of the folktale [6].

Propp's theory lays out a grammar in three levels, where the middle level, that of the *function*, has 31 pieces that describe the major constituents of Russian fairy tales. These pieces include such plot fragments as *Villainy*, *Struggle* (between the Hero and Villain), and *Reward* (of the Hero for defeating the Villain). I demonstrated that these plot pieces and rules of combination, rather than being figments of Propp's imagination, can be learned by a computerized correlation method from the actual tales [4]. I call the method *Analogical Story Merging*, which is modification of a machine learning technique called *Bayesian Model Merging* that relies on correlations uncovered by a statistical process leveraging Bayes' rule. Key to the method is a bias function called the *prior* that tells the method what similarities it should consider important when considering what parts of the folktales may be patterns. In the case of learning Propp's theory, the prior focuses the method on three important features: the semantic character of events; which characters are involved in those events; and where the events occur in the timeline of the tale.

This is all well and good. We have a method that extracts higher-level plot patterns from sets of folktales, where the patterns themselves cannot be seen by examining just a single tale. We have identified information flow from a set of tales, and in contrast to other information in an individual folktale, there is a fair chance that this information actually reflects something in reality (rather than a fictional world) — it likely reflects the ideas of participants in the culture under examination, such as the sorts of bad things that can happen to people, the proper conduct of a heroic person, and the rewards for heroic behavior. But is this the only information transmitted by sets of folktales?

Consider a competing proposal for narrative structure, that of Lévi-Strauss [5]. In his structural analysis of myth, he identified units of analysis that he called *mythemes*, where each mytheme was a set of semantic units unified by their treatment of a common theme, such as *death* or *familial relations*. In contrast to Propp's so-called *diachronic* analysis of the tales, where each function is laid out in the order it is encountered in the tale, Lévi-Strauss organized his analysis *synchronically*, where mythemes are organized by theme regardless of their position in the tale. Moreover, Lévi-Strauss's 'grammar' (if it may be called that) is highly constrained, consisting of two paired binary oppositions arranged in a specific relationship. While I don't have an algorithm that demonstrates learning Lévi-Strauss's theory from stories, it is clear that the method I used for learning Proppian structures would not be sufficient, merely from theoretical limitations of grammatical inference.

Given both Propp's and Lévi-Strauss's analyses, what are we to say about the information they contain, relative to one another? Lévi-Strauss's theory is not a specialization of Propp's, or vice-versa — they are completely orthogonal. One needs a completely different method to learn Lévi-Strauss's theory from the stories. So which is the actual information carried across the set? The answer is clear, in that it depends on the method one uses. Both theories, if underwritten by regularities in the distributed system (of people, culture and folktales) describe equally valid information carried by the set. They are two completely different interpretations of what is going on.

This observation points the way toward understanding what is going on with different degrees of intensionality. Indeed, these examples show that intensionality is not a matter of degrees at all. We have an intuitive ranking of the dog's, child's, and Dretske's classifications of the daisy, but these are all relative to an implicit value judgement about merely one aspect of the classification method. Dretske's classification is a refinement of the child's, and the child's a refinement of the dog's (we suppose), and we have implicitly associated a more refined classification with a higher degree of intensionality. But in the case of narrative structure, there is no such refinement relationship. The two theories attend to quite different patterns in the texts at hand. Thus we see clearly that intensionality, in the general case, does not come with a clear intuitive ranking. Intensionality is relative to some measure on the method we are using to extract our information. We may find, for a particular situation, that complexity of the method, or complexity of classification, or utility for some purpose, are the appropriate way to rank and order intensionalities. We have refined the question from *what makes this processor of information more intensional than that one?* to *what characteristics of the classification method explain our intuitions of degree of intensionality?*

5 Information in Individual Narratives

Information can flow from a set of narratives. What can we do with it? We can of course go back and apply that knowledge to a single narrative. For example, in a Proppian-style analysis, suppose we have learned from our set of narratives that there is something we decide to call a *Villainy* in the culture in question, and it takes certain specific forms. The method shows us what to pay attention to for when looking for villainies, and so now we can (usually) pick out a villainy in an individual story. This does not mean, of course, that the individual narrative contains the information about the nature of villainies — we learned that from the set of narratives. Our concept of villainy depended on analyzing the whole set. It contains the information that there is (or is not) a villainy in that particular narrative. This is just another way of emphasizing, as the DBS theory does, that information flow is relative to the receiver. Interestingly, for cultural narratives, what information flows at the narrative structure level is a function not only of the method used (e.g., a Proppian-style analysis, or Lévi-Straussian analysis, or something else), but also a function of the contents of the set itself. Change the set of tales, to folktales from another culture, and you get different functions [4, §6.1.5]. This naturally leads to the questions of how we decide what set of narratives to analyze? What principles should guide that selection? In my work, and Propp's, the principle was a representative selection of a particular genre of folktale from a particular culture. For other purposes the principle could be quite different.

There is a second point of interest. Naturally, even if one keeps the selection principle the same, the nature of the information extracted from the set of folktales varies with the number of tales in the set [4, Fig. 5-3]. For smaller sets, we learn fewer Proppian, and they are learned with less fidelity. Thus, in a sense, when fewer tales support the higher-level analysis, the information carried by the individual tale is coarser, and the information "chunk size" is larger. One would imagine, when doing a Proppian analysis on a set that properly contains the set of tales analyzed by Propp, that one might find more than 31 functions. (Indeed, I noted a possible missing function of this sort [4, §5.5.4].) In this case, with more tales, the information chunk is smaller, and the information carried by the tale is finer.

In a Proppian-style analysis, if Propp's functions can be considered the "information unit" or "chunk size" of the plot of the narrative, relative to some particular set of folktales, how do we know when we have the right sized chunk? I cannot think of any philosophical reason that the chunks, in this particular case, will be of one size rather than another. I imagine it will boil down to an experimental question, where one may find, upon adding more and more folktales to the original set, that the chunk size does not get any smaller. For another style of analysis one may find that there is no stable point, and the chunk size always depends on the number of narratives added. This could potentially be a discriminator between effective and ineffective theories of narrative structure.

Acknowledgements. This work was funded by the Office of Naval Research under award number N00014-09-1-0597. Any opinions, findings, and conclusions or recommendations expressed here are those of the author and do not necessarily reflect the views of the Office of Naval Research.

References

1. Barwise, J., Seligman, J.: Information Flow: The Logic of Distributed Systems. Cambridge University Press, Cambridge (1997)
2. Dretske, F.I.: Knowledge and the Flow of Information. MIT Press, Cambridge (1981)
3. Dretske, F.I.: Précis of knowledge and the flow of information. The Behavioral and Brain Sciences 6(1), 55–90 (1983)
4. Finlayson, M.A.: Learning Narrative Structure from Annotated Folktales. Ph.D. thesis, Department of Electrical Engineering and Computer Science. Massachusetts Institute of Technology (2011)
5. Lévi-Strauss, C.: The structural study of myth. The Journal of American Folklore 68(270), 428–444 (1955)
6. Propp, V.: Morphology of the Folktale. University of Texas Press, Austin (1968)