<u>ESSLLI 2001, Helsinki, Finland</u>

The Mathematics of Information

Lecture 1: Shannon's Information Theory

Keith Devlin*

Claude Shannon

Mathematician Claude Shannon was born in Michigan in 1916 and died in 2001. The approach to information and communication he laid out in his groundbreaking paper *A Mathematical Theory of Communication*, published in the Bell System Technical Journal in 1948, and republished virtually unchanged in the pamphlet *The Mathematical Theory of Communication* he wrote with Warren Weaver the following year (published by the University of Illinois Press) remain current to this day. (Note how the "a" of his paper became "the" in the Shannon-Weaver version.)

After obtaining degrees in both mathematics and engineering at the University of Michigan, Shannon went to MIT to pursue graduate studies in mathematics. There he came into contact with some of the men who were laying much of the groundwork for the information revolution that would take off after the Second World War, notably the mathematician Norbert Wiener (who later coined the term cybernetics for some of the work in information theory that he, Shannon, and others did at MIT and elsewhere) and Vannevar Bush, the dean of engineering at MIT (whose conceptual "Memex" machine foretold the modern World Wide Web and whose subsequent achievements included the establishment of the National Science Foundation).

In the early 1930s, Bush had built a mechanical, analog computer at MIT called the Differential Analyzer, designed to solve equations that were too complex for the (mechanical) calculating machines of the time. This massive assemblage of cog wheels, shafts, gears, and axles took up several hundred feet of floor space, and was powered by electric motors. Preparing the device to work on a particular problem required physically configuring the machine, and could take two or three days. After the machine had completed the cycle that constituted "solving" the equation, the answer was read off by measuring the changes in position of various components.

Always a tinkerer, Shannon took to working with the Analyzer with great enthusiasm. At Bush's suggestion, for his master's thesis, he carried out a mathematical analysis of the operation of the machine's relay circuits. In 1938, he published the results of this study in the Transactions of the American Institute of Electrical Engineers under the title "A Symbolic Analysis of Relay and Switching Circuits." Bush's seemingly mundane motivation for having Shannon do the work was the telephone industry's need for a mathematical framework in which to describe the behavior of the increasingly complex automatic switching circuits that

*Center for the Study of Language and Information, Stanford University, Stanford, California 94305. devlin@csli.stanford.edu

were starting to replace human telephone operators. What Shannon produced far transcended that aim. The ten page article that he published in the *Transactions of the AIEE* has been described as one of the most important engineering papers ever written. And with good reason: quite simply, it set the stage for digital electronics.

Shannon began by noting that, although the Analyzer computed in an analog fashion, its behavior at any time was governed by the positions of the relay switches, and they were always in one of just two states: open or closed (or on or off). This led him to recall the work of the nineteenth century logician George Boole, whose mathematical analysis of the "laws of thought" was carried out using an algebra in which the variables have just the two "truth values" T and F (or 1 and 0). From there it was a single — but major — step to thinking of using relay circuits to build a digital "logic machine" that could carry out not just numerical computations but also other kinds of "information processing."

In 1940, Shannon obtained his doctorate in mathematics, and went to the Institute for Advanced Study at Princeton as a National Research Fellow, where he worked with Hermann Weyl. The following year, he took a position at the Bell Telephone Laboratories in New Jersey, joining a research group who were trying to develop more efficient ways of transmitting information and improving the reliability of long-distance telephone and telegraph lines.

Prior to Shannon's work, mathematicians and engineers working on communications technology saw their job as finding ways to maintain the integrity of an analog signal traveling along a wire as a fluctuating electric current or through the air as a modulated radio wave. Shannon took a very different approach. He viewed information as being completely encoded in digital form, as a sequence of 0s and 1s — which he referred to as "bits" (for "binary digits"), following a suggestion of his Princeton colleague John Tukey.

In addition to providing the communications engineers with a very different way of designing transmission circuits, this shift in focus also led to a concept of information as an objective commodity, disembodied from a human sender or receiver. After Shannon, the name of the game became: how can you best send a sequence of discrete electrical or electromagnetic pulses from one point to another?

A particular consequence of this new approach, which Shannon himself was not slow to observe, was that whereas even a small variation in an analog signal distorts — and can conceivably corrupt — the information being carried by that signal, the discrete yes-or-no/on-or-off nature of a digital signal means that information conveyed digitally is far less prone to corruption; indeed, by adding extra bits to the signal, automatic error detection and correction can be built into the system. (A feature of digital coding that is exemplified by the oft-repeated claim of CD manufacturers that you can drill a centimeter hole in your favorite music CD and it will still play perfectly.)

From a mathematical point of view, arguably the most significant aspect of Shannon's new, digital conception of information is that it provides a way to measure information. This is the topic of this lecture.

Shannon's communication model

Figure 1 gives Shannon's basic communication model, consisting of an information source (the transmitter), the channel over which the information is transmitted, and the receiver. Also indicated is the fact that, during transmission, the message may be subject to distortion, due to noise on the channel.

Figure 2 is a more detailed model of communication, showing the various components of a

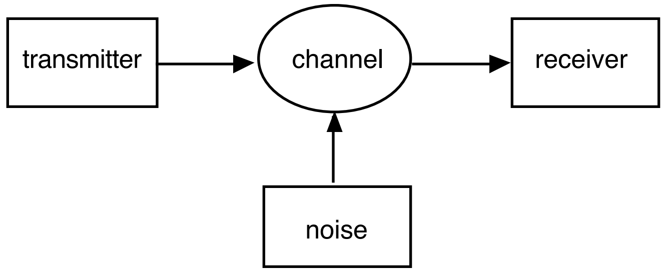


Figure 1: Shannon's basic communication model

modern electronic communication system:

- *information source:* The information is in the form of discrete data (usually digitally represented).
- *data reduction:* Source data not relevant to the destination is removed. The remaining information is called the *effective information*.
- *data compression:* Also called *source coding*.
- *encryption*.
- *channel coding:* Information is added to increase the likelihood that the original message can be successfully reconstructed even if the signal is distorted during transmission due to noise.
- *modulation:* The data is put into a form suitable for transmission over a continuous medium — usually by modulating an electrical current or electromagnetic wave.
- *transmission:* Transmission is inevitably accompanied by distortion of the signal due to noise.
- *demodulation:* The continuous (modulated) signal is converted back into discrete (digital) information.
- *channel decoding:* Any errors introduced during transmission are (hopefully) detected and corrected.
- *decryption*.
- *data decompression*. Also called *source decoding*.
- *data reconstruction:* The data is put into the desired form for the destination.

Shannon's procedure for measuring information assumes a communication system of the kind just described. To present his theory, we need some elementary notions from probability theory.

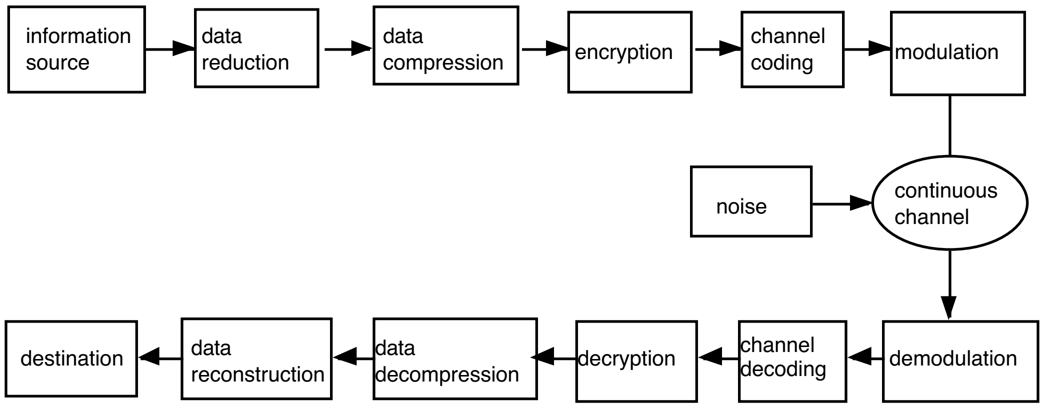


Figure 2: More detailed communication model

Probability theory

Consider a *probabilistic experiment* $\mathcal{E}$ with possible outcomes (i.e., *events*) $x_1, \ldots, x_n$. The set $X = \{x_1, \ldots, x_n\}$ is called the *outcome space* for the experiment. The probability of the event x_i occurring is $p(x_i)$, or simply p_i. The set $P = \{p_1, \ldots, p_n\}$ is called the *probability distribution*.

For example, if $\mathcal{E}$ is the experiment of throwing a die, the outcome space would be $X = \{x_1, \ldots, x_6\}$, where x_i is the event of throwing an i for each $i = 1, \ldots, 6$. If the die is honest, then $p_1 = \ldots = p_6 = 1/6$.

Any probability distribution satisfies two basic requirements:

- $p_i \geq 0$, for all i;
- $\displaystyle\sum_{i=1}^{n} p_i = 1$.

Sometimes an experiment has two kinds of outcome. For example, if we are sampling electronic components coming off a production line, the sampling can provide a batch number for the component tested and a quality rating for that component. In such a case, the experiment really consists of two subexperiments, $\mathcal{E}$ and $\mathcal{F}$, say, with sample spaces $X = \{x_1, \ldots, x_n\}$ and $Y = \{y_1, \ldots, y_m\}$ and associated probability distributions $P = \{p_1, \ldots, p_n\}$ and $Q = \{q_1, \ldots, q_m\}$, respectively. We can view $(\mathcal{E}, \mathcal{F})$ as an experiment whose outcomes are pairs (x_i, y_j) with $x_i \in X$ and $y_j \in Y$. Let $r_{ij} = p(x_i, y_j)$ be the probability that event (x_i, y_j) occurs. This is called the *joint probability*. Given the joint probability, one can derive the probabilities p_i and q_j, which are then called the *marginal probabilities*. It is easy to see that

$$
\begin{aligned}
p_i &= \sum_{j=1}^{m} r_{ij}, \quad \text{for } i = 1, \ldots n \\
q_j &= \sum_{i=1}^{n} r_{ij}, \quad \text{for } j = 1, \ldots m
\end{aligned}
$$

Since $\displaystyle\sum_{i=1}^{n} p_i = 1$ and $\displaystyle\sum_{j=1}^{m} q_j = 1$, it follows that

$$\sum_{i=1}^{n}\sum_{j=1}^{m} r_{ij} = 1$$

Another possibility is when an experiment $\mathcal{E}$ is dependent on an experiment $\mathcal{F}$. The probability that event x_i will occur, given that event y_j has occurred, is called the *conditional probability*, and is denoted by $p(x_i|y_j)$. For example, a natural language interpreter might want to know the probability that "N" will be the next input letter given that the input so far is the string "INFORMATIO".

The conditional probability of x_i given y_j is defined by:

$$p(x_i|y_j) = \frac{r_{ij}}{q_j} = \frac{p(x_i, y_j)}{p(y_j)}$$

provided $q_j > 0$. Similarly,

$$p(y_j|x_i) = \frac{r_{ij}}{p_i} = \frac{p(x_i, y_j)}{p(x_i)}$$

provided $p_i > 0$.

The above equations can be rewritten as:

$$p(x_i, y_j) = p(y_j)p(x_i|y_j) = p(x_i)p(y_j|x_i)$$

This can be rearranged to give Bayes' theorem that, whenever $p(y_j) > 0$:

$$p(x_i|y_j) = \frac{p(x_i)p(y_j|x_i)}{p(y_j)}$$

This enables $p(x_i|y_j)$ to be calculated when $p(y_j|x_i)$ is known. The formula can be rewritten as:

$$p(x_i|y_j) = \frac{p(x_i)p(y_j|x_i)}{\sum_{i=1}^{n} p(x_i)p(y_j|x_i)}$$

Notice that, for any given y_j,

$$\sum_{i=1}^{n} p(x_i|y_j) = 1$$

since one of $x_1, \ldots, x_n$ must occur.

If $p(x_i|y_j) = p(x_i)$ (i.e., the occurrence of y_j has no influence on the occurrence of x_i), then $p(x_i, y_j) = p(x_i)p(y_j)$ and $p(y_j|x_i) = p(y_j)$. In this case we say that the events are *independent* of each other. The reverse is also true: if $p(x_i, y_j) = p(x_i)p(y_j)$, then $p(y_j|x_i) = p(y_j)$ and $p(x_i|y_j) = p(x_i)$.

Two experiments $\mathcal{E}$ and $\mathcal{F}$ are called *statistically independent* if $p(x_i, y_j) = p(x_i)p(y_j)$ for all i, j.

An experiment $\mathcal{E}$ is said to be *completely dependent* on an experiment $\mathcal{F}$ if, for all j there is a unique i, say i_j, such that $p(x_{i_j}|y_j) = 1$ (equivalently, $p(x_{i_j}, y_j) = p(y_j)$).

Shannon's information measure

Suppose that eight people work in an office, four men, four women. The eight employees meet to decide among themselves who will represent them in some upcoming contract negotiations. They make their decision and convey it to their supervisor, who sends the boss an email informing her who has been chosen. The boss views all eight as equally qualified for the role and will accept whichever name comes forward. Before the employees make their decision there are eight equally likely outcomes. The employees' decision reduces that eight to the one person selected. In quantitative terms, what is the information content of the email the boss receives from the supervisor?

Following Shannon, we can answer this question as follows. Suppose that two of the men are tall, two short, and similarly for the women. Suppose further that all of the employees started to work in the office at different times. Then it is possible to identify any one of the 8 employees by giving 3 items of binary information: man or woman, tall or short, more recent hire. The first item of information (man or woman) reduces the 8 possibilities to 4, the second item of information (tall or short) reduces those 4 to 2, and the last item (more recent hire) picks one of those two. We can encode these criteria using 0 and 1 as follows: man = 0, woman = 1; tall = 0, short = 1; newer = 0, older = 1. Then the email could consist of a single binary sequence of length 3, say 001 (the longer serving of the two tall men) or 110 (the shorter serving of the two short women). According to Shannon, such a message will carry 3 *bits* (for binary digits) of information. (It's not hard to see that the identity of the employee chosen could not be specified with fewer than 3 bits, no matter what criteria were used to subdivide the employees.)

The general idea then is to measure the information content of a message $\mathcal{E}$ by the (least) number of bits required to convey the information. If $\mathcal{E}$ reduces n equally likely possibilities to 1, then the (Shannon) information content of $\mathcal{E}$ is

$$I(\mathcal{E}) = \log n$$

where the logarithm is taken to base 2. In the case of the eight office employees, where $\mathcal{E}$ is the supervisor's email message, $I(\mathcal{E}) = \log 8 = 3$ (because $2^3 = 8$). If there had been ten employees in the office, the information content of the email would have been $\log 10 \approx 3.3$ bits. Of course, with digital communication technologies you can't transmit a fractional number of bits, so in this case it would require 4 actual bits to convey the message. But the information content, as a mathematical concept, is (to one decimal place) 3.3 bits.

So far, we have restricted attention to the case where a message picks out one of a number of equally likely possibilities. Now let's consider the more general case where the possible outcomes, say $x_1, \ldots, x_n$, are not all equally likely. Let $p_i = p(x_i)$ be the probability that x_i occurs. The greater p_i is, the more we expect x_i to occur, and the less will be the information content of the message that tells us x_i has in fact occurred. In particular, if $p_i = 1$, then it is certain that x_i will occur, and the message tells us nothing we did not already know — its information content is 0. In general, the information content of the message s_i will be

$$I(x_i) = \log 1/p(x_i) = \log 1/p_i$$

In the case where all outcomes are equally likely, we have $p_1 = \ldots = p_n = 1/n$, of course, so this reduces to the definition we had before.

Notice that we can also write the formula as:

$$I(x_i) = -\log p(x_i) = -\log p_i$$

The quantity $I(x_i)$ is sometimes called the *surprisal value* of the event x_i.

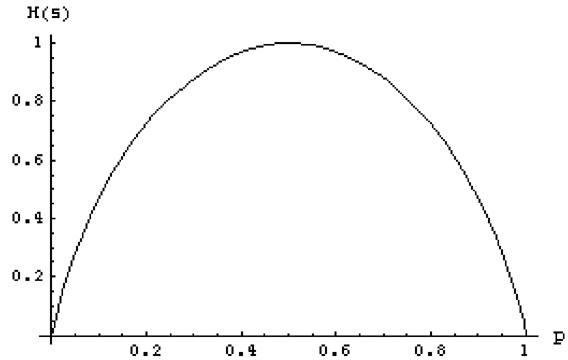


Figure 3: The average information generated by tossing a biased coin with a probability p of coming up heads. The information content is greatest for an honest coin, where $p = 0.5$.

For example, suppose we flip a biased coin for which the probability of throwing heads on any given throw is 0.9. Then the amount of information associated with throwing a head is $\log 1/.9 = .15$ bits and the amount of information associated with throwing a tail is $\log 1/.1 = 3.33$ bits.

In many cases, in particular in communications engineering, we are not concerned with the information content of a particular signal but the average information generated by a given source $\mathcal{E}$. This is given by the formula

$$H(\mathcal{E}) = \sum_i p(x_i)I(x_i)$$

(The letter H is used in honor of R.V.L. Hartley, who first tried to find a quantitative measure of information in 1928, before Shannon did his work.) That is, take the information content $I(x_i)$ of each possibility from the source, weighted according to the probability $p(x_i)$ of its occurrence.

In the case where $\mathcal{E}$ is the tossing of the biased coin, this gives

$$\begin{aligned} H(\mathcal{E}) &= p(\textsc{head})I(\textsc{head}) + p(\textsc{tail})I(\textsc{tail}) \\ &= (0.9)(0.15) + (0.1)(3.33) \\ &= 0.467 \text{ bits} \end{aligned}$$

If an honest coin is tossed, $I(\textsc{head}) = I(\textsc{tail}) = \log 1/.5 = 1$ and the average information of the outcomes is

$$\begin{aligned} H(\mathcal{E}) &= p(\textsc{head})I(\textsc{head}) + p(\textsc{tail})I(\textsc{tail}) \\ &= (0.5)(1) + (0.5)(1) \\ &= 1 \text{ bit} \end{aligned}$$

Thus, less information is generated by flipping a biased coin than flipping an honest coin. Figure 3 shows the graph of the variation of $H(\mathcal{E})$ as p goes from 0 to 1.

By way of an additional example, consider a modification to the original office scenario. This time, the supervisor judges Sally to be unsuitable for the job. He decides in advance that if the employees select Sally, then he will send in Anne's name instead. What is the information content of the supervisor's email in this case?

The probability that Anne's name will be in the email is $2/8 = 0.25$, and hence the information content in that case will be

$$I(\text{Anne}) = \log(1/.25) = \log 4 = 2 \text{ bits}$$

The probability that Sally's name is in the email is 0, of course. And for each other employee A, $p(A) = .125$ and $I(A) = \log(1/.125) = \log 8 = 3$ bits. Thus, the average information content of the email, e, is

$$H(e) \;=\; \sum_A p(A)I(A) \;=\; [.25 \times 2] \;+\; [0] \;+\; 6 \times [.125 \times 3] \;=\; 2.75 \text{ bits}$$

This is less than the information content of 3 bits in the original scenario.

In general, then, if $\mathcal{E}$ is a probabilistic experiment with outcome space $X = \{x_1, \ldots, x_n\}$ and probability distribution p, and if we write $p_i = p(x_i)$, then the average amount of information in performing $\mathcal{E}$ is:

$$H(\mathcal{E}) = -\sum_{i=1}^{n} p_i \log p_i$$

Other common notations for the Shannon information measure are $H(X)$, $H(P)$, and $H(p_1, \ldots, p_n)$.

Lemma

1. $H(P)$ is a continuous function of p.
2. $H(P)$ is symmetric: The value of $H(p_1, \ldots, p_n)$ does not depend on the ordering of $p_1, \ldots, p_n$.
3. $H(P)$ is additive: If X and Y are two sample spaces, where the outcomes in X are independent of those in Y, then for the information relating to joint events (x_i, y_j),

$$H(p_1q_1, \ldots, p_1q_m, \ldots, p_nq_1, \ldots, p_nq_m) = H(p_1, \ldots, p_n) + H(q_1, \ldots, q_m)$$

4. $H(P)$ is maximum if all probabilities are equal; $H(P)$ is minimum if one outcome has probability 1.
5. $H(P) \leq \log n$, with equality if and only if $p_i = \dfrac{1}{n}$ for all $i = 1, \ldots, n$.

Proof: Only the last statement takes a bit of effort. We have:

$$\begin{aligned}
H(P) - \log n \quad &= \quad -\sum_{i=1}^{n} p_i \log p_i - \log n \\
&= \quad -\sum_{i=1}^{n} p_i [\log p_i + \log n] \qquad (\text{since } \sum_{i=1}^{n} p_i = 1) \\
&= \quad \sum_{i=1}^{n} p_i \log \frac{1}{p_i n}
\end{aligned}$$

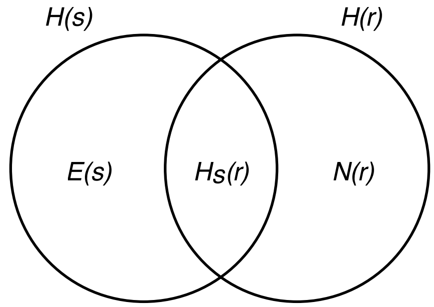


Figure 4: Equivocation, noise, and information transmitted

Now, for any $a > 0$, $\ln a \leq a - 1$, so

$$\log a = \frac{\ln a}{\ln 2} \leq \frac{a-1}{\ln 2} = (a-1)\frac{\ln e}{\ln 2} = (a-1)\log e$$

Using this inequality, we get, from our earlier result:

$$\begin{aligned} H(P) - \log n &\leq \sum_{i=1}^{n} p_i \left[\frac{1}{p_i n} - 1\right] \log e \\ &= \left[\sum_{i=1}^{n} \frac{1}{n} - \sum_{i=1}^{n} p_i\right] \log e \\ &= \left[n\frac{1}{n} - 1\right] \log e \\ &= 0 \end{aligned}$$

Hence $H(P) \leq \log n$, with equality if and only if $\frac{1}{p_i n} = 1$, which is what we set out to prove. □

Transmission of information

Consider now the case where information is transmitted from a source s to a receiver r. The average amount of information generated at s is $H(s)$. Let $H_s(r)$ be the average amount of information about s received at r.

Clearly, $H_s(r) \leq H(r)$; the information received at r about s cannot exceed the information received at r. But equality is not guaranteed. For example, in the case of the eight office employees, suppose that the original scenario is modified in that the supervisor sends the name of the chosen representative not by email but by handing a scribbled note to a courier. Then the information content at the source, s, is $H(s) = 3$ bits. However, the courier loses the note, but rather than admit his error, he simply writes down one of the employees' names at random, say "Harry". Since the name "Harry" is one of eight equally likely names that the boss could have received, $H(r) = 3$ bits. However, the note the boss reads actually contains *no* information about what actually took place at the employees' meeting, so $H_s(r) = 0$. This is the case, even if, as it happens, the employees had chosen Harry as their representative. Referring to Shannon's transmission model (Figure 1 or 2), $H(r)$ consists entirely of *noise* on the channel. As far as the communication process is concerned, the information generated at the source, $H(s)$, is all *equivocation*.

The general situation is illustrated by the Venn diagram in Figure 4. $H_s(r)$ is the region common to both circles — the information generated at s that is actually received at r. The region $E(s)$ is the equivocation at the source — the information generated at s that is lost in the transmission process. The region $N(r)$ is the noise — the information received at r that was not generated at s. Symbolically:

$$H_s(r) \;=\; H(r) - N(r) \;=\; H(s) - E(s)$$

To assign numerical values to the quantities $N(r)$ and $E(s)$, suppose that the outcome space for s is $\{x_1, \ldots, x_n\}$.

Suppose x_i is generated at s. In the absence of noise and equivocation, x_i will be received at r. But if there is noise or equivocation, then it is possible that some other signal, say x_j, arrives at r. The probability of this happening is $p(x_j|x_i)$. Thus, the information content (i.e. surprisal value) of the arriving signal being x_j is $\log 1/p(x_j|x_i) = -\log p(x_j|x_i)$. To calculate the total noise associated with transmission of the signal x_i, we sum all of these individual noise terms, each one weighted by the probability that is occurs, namely:

$$N(x_i) = -\sum_{j=1}^{n} p(x_j|x_i) \log\, p(x_j|x_i)$$

(Note that the above sum includes $j = i$. We are measuring the spread of all received signals for a given source signal. As we saw in our example above, it is possible to receive the “right” signal by accident.) The average noise associated with the transmission to r is obtained by taking the weighted sum over all possible source signals:

$$N(r) = \sum_{i=1}^{n} p(x_i) N(x_i)$$

Similarly, turning our attention to the receiver, r, suppose x_i is received at r. The equivocation associated with the receipt of the signal x_i is

$$E(x_i) = -\sum_{j=1}^{n} p(x_i|x_j) \log\, p(x_i|x_j)$$

and the average equivocation associated with the source s is

$$E(s) = \sum_{i=1}^{n} p(x_i) E(x_i)$$

Shannon defines the *capacity* of an information channel as the maximum transmission rate in bits per second. I state but do not prove Shannon’s fundamental theorem on transmission rate:

Theorem (Shannon’s Coding Theorem)

Given a noisy channel with capacity C (in bits per second) and an amount of information H, if $H \leq C$ then it is possible to transmit the information over the channel with arbitrarily small probability of error.

If $H > C$, it is possible to encode the source so that the equivocation is less than $H - C + \epsilon$, where $\epsilon > 0$ is arbitrarily small. But there is no coding method that gives an equivocation less than $H - C$. $\square$

[Note: The proof of Shannon’s theorem is nonconstructive. It does not provide any actual codings.]

Limitations of Shannon's theory

Shannon developed his theory in order to study the behavior of communication systems in a quantitative manner. The theory serves that purpose well. But it has several limitations that prevent its use beyond Shannon's original aims.

First, it concentrates on the *average* performance of a communication channel, not what happens on any particular occasion. For example, suppose I wish to transmit to you the position of a king on a chessboard, where the squares are numbered 1 through 64 and the king is placed on the board at random. It requires 6 bits to identify the king's position. Suppose however that the only channel available to me has a capacity of one bit per second, and my message is limited to one second duration. Then it would seem that I cannot send the required message. Suppose, however, that I set up a protocol whereby transmission of a 1 means the king is on square 2 and transmission of a 0 means it is not on square 2. Then, on each occasion when the king is placed, my message will convey correct information. Of course, most of the time my message will not tell you where the king is, rather just that it is not on square 2. But if, by chance, the king is on square 2, I will be able to inform you of that fact. On such an occasion, the channel does allow me to transmit a message with 6 bits of information. The limitation of 1 bit per second is a bound on the *average* amount of information that can be transmitted. It does not prevent the transmission of more information on a particular occasion. But it is easy to imagine scenarios where it is what happens on a particular occasion that is important, not the average behavior.

Second, Shannon's approach only works in the case where the messages under consideration serve to select from some a priori set of possibilities, whose probability distribution is known. If no such prior distribution is known, the theory is of little help.

Moreover, the theory says nothing about the informational content of any particular message — what does the message say? For example, according to Shannon's measure, the message that says a person tossing a coin has just thrown a head has the same information content as the message that tells you which of two equally matched teams has just won the Superbowl Final. In each case the message reduces two equally probable possibilities to one. Yet the two items of information are entirely different, and by almost any reasonable, everyday measure, the latter is more informative (i.e., carries more information). Again, by Shannon's metric, my announcement that I just saw a leopard in my garden has greater information content than my message that I just saw a dog in my garden. True enough, the former has greater surprisal value. But in a natural sense they carry the same amount of information: namely, the fact that I saw an animal in my garden plus identification of that animal.

Thus, it is a historical misnomer to call Shannon's framework a theory of information, as many have done. It is not; it is a theory of communication channels and of channel capacity. Nevertheless, as Dretske observed in the 1980s, it is possible to use elements of Shannon's theory to develop a qualitative theory of information content. This is the topic of the second lecture.

Reference

Van der Lubbe, J.C.A., *Information Theory*, UK: Cambridge University Press (1997).

<u>ESSLLI 2001, Helsinki, Finland</u>

The Mathematics of Information

Lecture 2: Dretske's Theory of Information

Keith Devlin*

Shannon's theory, although it is sometimes referred to as a "theory of information," is nothing of the sort. It is a theory of coding and of channel capacity. To achieve his desired aim, Shannon concentrated on the average amount of data generated by a source or transmitted through a channel. What that data represented was irrelevant to his purposes. Indeed, whereas it is possible to average numerical amounts of data, there is no such thing as an "average" of information content.

For instance, what is the average of the information that this article is written in TEXand the information that we are currently in Finland? Information — the content of the signals Shannon studied — is something associated with a *particular* signal. Averaging the number of bits it takes to represent these two pieces of information bears no relationship to the information itself.

In fact, in most everyday situations, it is impossible to assign a unique bit count to an item of information, since there is no preferred range of alternative possibilities. For example, against what range of possibilities do you measure the information that I had steak for dinner last night?

To develop a genuine theory of *information*, we need to look at how a particular signal carries information. And in order to do that, we have to ask ourselves, what exactly do we mean by information?[1]

The nature of information

Whatever information is — and this course is intended to provide some kind of answer — people and other cognitive agents — various kinds of organism, certain mechanical or material devices — use it. They make their way in the world by acquiring information and reacting accordingly. A human being, on seeing flames all around, will take flight, having extracted the information that there is a fire and being already in possession of the previously acquired information that fires are life-threatening; a cat, on seeing a mouse, will pounce, knowing that there is fun to be had, if not food; a flower, on detecting the sun's rays in the morning, will open; a thermostat, on sensing a drop in temperature, will switch on the heating; and so on.

Already these few examples indicate one significant factor that we shall have to consider right at the outset: the fact that different agents are capable of extracting different information

*Center for the Study of Language and Information, Stanford University, Stanford, California 94305. devlin@csli.stanford.edu

[1]Most of the following section is taken from my book [1].

from the same source. A person can pick up a great deal of information about the surrounding air — its cleanliness, the presence or absence of any smells, whether it is stationary or moving, whether it is warm or cool, how humid it is, and so on. A simple, mechanical thermostat on the other hand can only pick up one piece of information, though it is a piece of information that few humans are capable of picking up in all circumstances: namely the information whether the temperature is above or below the value set, say 21 °C. Just what kind of information may be picked up depends upon just what kind of device the agent is, and in particular upon the *state* of that agent *vis á vis* the extraction of information.

For a richer example, suppose you come across a tree stump in the forest. What information can you pick up from your find? Well, if you are aware of the relationship between the number of rings in a tree trunk and the age of the tree, the stump can provide you with the age of the tree when it was felled. To someone able to recognize various kinds of bark, the stump can provide the information as to what type of tree it was, its probable height, shape, leaf pattern, and so on. To someone else it could yield information about the weather the night before, or the kinds of insects or animals that live in the vicinity; and so on. (As an exercise, you might like to think of, say, three further items of information the tree stump may be capable of providing.)

What the above examples indicate above all is that the acquisition of information from a source depends upon various relationships[2] of which the agent is aware, or to which the agent is attuned. Information is obtained from a source by means of some relationship. Thus the simple thermostat is, by its construction, attuned to a certain physical relationship between temperature and the expansion rates of various metals, and the forest ranger is aware of the relationship between the number of rings in a tree trunk and the age of the tree. (But note that this is not to say that the thermostat is in any sense 'aware' of how and why it operates the way it does. Nor does the forest ranger need to have a deep knowledge of tree biology. Attunement to, or behavior-guiding awareness of, a particular relationship that gives rise to information may be an essentially *empirical* phenomenon.)

The example of the tree stump also highlights another aspect of information acquisition: there is never just a single item of information to be obtained from a given situation. For one thing, different relationships will give rise to different (kinds of) information being picked up. In addition, there is the phenomenon of 'nested' information. Given the possession of prior information and/or attunement to other information-yielding relationships, the acquisition by an agent of an item of information Φ can also provide the agent with an additional item of information Ψ. For instance, if I tell you that a certain geometric figure (hidden from your view) is a square, then you also gain the (different) information that the figure is a rectangle. (This is by virtue of the 'analytic' relationship that all squares are rectangles.) You also learn that the figure is a parallelogram, that all its internal angles are 90 degrees, that it is a planar figure, and so on. Each of these additional items of information is *nested* (by way of certain relationships) in the information that the figure is a square. Thus information, when it comes, tends to do so in large, structured amounts. The acquisition of certain specific items of information from the mass available is part of *cognition.*

Information, as we usually encounter it, is not unlike a bottomless pit, seemingly capable of further and further penetration. On the other hand, cognitive agents deal (at any one moment) with a relatively small collection of specific *items* of information extracted from that fractal-like mass of available information. The acquisition of information from the environment by a cognitive agent is a process analogous to, though not necessarily the same as, going from the infinite and continuous to the finite and discrete. Dretske used the terms *analog* and *digital*

[2]When we come to develop a situation-theoretic account of information, we shall use the term 'constraint' to refer to relationships that yield information.

coding to facilitate discussion of this process.

Consider the most basic kind of informational item: the possession or otherwise by an object s of a property P. Such an item of information will be conveyed to an agent by means of some kind of signal. That signal is said to carry the information that the object s has the property P in *digital* form, if that signal carries no additional information about s other than that which is nested in s being P. Otherwise the signal is said to carry the information that s is P in *analog* form.

Thus if I tell you that John is taller than Fred, that piece of information is conveyed to you in digital form. You learn nothing more about John and Fred other than that one single fact (plus anything nested therein, such as John's being able to reach a higher shelf in the library than Fred, and so on). You do not, for instance, pick up the information that John has a beard and Fred does not. If, on the other hand, I show you a photograph of John and Fred standing side by side, that photograph will also convey to you the information that John is taller than Fred, but it does so in analog form. By looking at the photograph, you are able to pick up all kinds of information about these two characters, information more specific than the mere fact as to who is the taller. For instance, you learn something about the *extent* by which John is taller than Fred, something that my simple utterance does not convey to you. Information received in analog form is always capable of yielding more specific details. Information received in digital form is not capable of such refinement. (It should be noted that all of this is relative to the item of information concerned. The photograph conveys in analog form the information that John is taller than Fred, but it stores in digital form the information of what the photographer saw through the camera lens at the time the picture was taken.)

We may regard the extraction of information from the environment by an agent as taking place in two separate stages, corresponding to the analog/digital distinction. The first stage is *perception*, where the information in the environment becomes directly accessible to the agent by way of some form of sensor (seeing, feeling, smelling, hearing, etc. in the case of a person or animal, or by the molecular agitation of the air impinging on the bimetallic strip in the case of a thermostat). At this stage the information flow is analog, relative to whatever information we are concerned with. The second stage, if there is one, involves the extraction of a specific item (or items) of information from that perceived continuum; that is to say, it involves the conversion from analog to digital information. This stage is *cognition.* In the case of a person, this could be the recognition from the photograph of the fact that John is taller than Fred; for a thermostat, it is the discrete behavior between open and closed. A *cognitive agent* is an agent that has the capacity of *cognition* in this sense; i.e., the ability to make the analog to digital conversion.

It should be emphasized that the last sentence above is not intended to constitute a *definition* of a cognitive agent. Rather it enunciates one of the attributes an agent must possess in order to justify the description *cognitive.* Dretske [2, p.142] says:

> It is the successful conversion of information into (appropriate) digital form that constitutes the essence of cognitive activity ... Cognitive activity is the *conceptual* mobilization of incoming information, and this conceptual treatment is fundamentally a matter of ignoring differences (as irrelevant to an underlying sameness) ... It is, in short, a matter of making the analog–digital conversion.

This much would appear to support use of analog–digital conversion as a working definition of 'cognition'. But in a footnote to the above quoted passage, Dretske goes on to say:

> It is not *merely* the conversion of information from analog to digital form that qualifies a system as a perceptual–cognitive system ... To qualify as a genuine per-

> ceptual system, it is necessary that there *be* a digital-conversion unit in which the information can be given a cognitive embodiment, but the cognitive embodiment of information is not *simply* a matter of digitalization.

Basically, the point Dretske is making is that, *provided* the agent has the means of manipulating and utilizing the information it obtains, *then* the digitalization of perceived information is the essence of cognitive activity [3].

Notice that the process of cognition (i.e., at the very least an analog–digital conversion) involves a loss (often a huge loss) of information. But this is compensated for by a very definite gain, in that there occurs a *classification* of the perceived information. A thermometer simply registers the temperature (analog coding of information) and hence is a perceiving agent but not a cognitive agent (in even the most minimal sense suggested above); the thermostat classifies the temperature into two classes ('warm' and 'cold'), and thus exhibits a form of cognition.

A suggestive term to use in this discussion would be *concept*. The thermometer has no *concept* of warm and cold, the thermostat does. It is by the use of concepts to classify perceived (i.e., incoming) information that such information becomes available for (semantic)[3] processing.

Consider an agent, $\mathcal{A}$, that is able to extract information from the environment, that is, can make the basic analog to digital conversion of perceived information. What cognitive abilities does the agent need in this regard?

One of the most important (and fundamental) cognitive abilities is the facility to *individuate* objects, that is to say, to see objects *as objects*. For example, when you see a table before you, you recognize it as *a table*, that is, as a single object, not as some huge collection of molecules or (on a different scale) an assembly of wood, steel, glue, and whatever; when I look at the computer screen in front of me as I type these words, I see (*recognize*) a single object, a computer screen; and so on. Of course, this is not to say that you or I are at all unaware that these individual objects are made up of many other, smaller objects. To you as you eat your lunch, the table is a single object. To the worker who made it, the single, individuated objects that mattered were not only the table, but also its various components, the legs, the top, the screws that hold it together, and so forth. And though I never regard my computer as anything other than a single, incomprehensible entity, I realize that to the computer engineer it has a fascinating and complex structure, involving many component objects.

It is, by and large, a matter of the agent's purpose (and to some extent of scale) just what parts of the world are individuated as single objects. The point is that in order to make their way in the world, cognitive agents make constant use of the ability to individuate or discriminate objects.[4]

In this study, entities that are individuated (by the agent or the theorist) as 'objects' will

[3]The parenthesized adjective is to distinguish between the kind of content-oriented manipulation usually referred to as 'information processing', which is what is meant here, as opposed to some kind of quantitative signal manipulation such as volume or frequency control.

[4]Individuation is perhaps unduly demanding. The less restrictive notion of *discrimination* is surely sufficient. To explain what this means, consider my dog, Sam. As a result of observing Sam's behavior, it seems reasonable to assume he has a number of individuation capacities; for instance he appears to individuate balls, sticks, our two cats, his water bowl, and many other objects. Far less clear is whether he *individuates* our house or the local wood. And yet his behavior varies in a systematic way according to whether he is in the house, outside the house in the garden, or in the local wood. That is to say, he *discriminates* (by his behavior) the house, the garden, and the wood. Thus in describing the dog's cognitive behavior, it seems appropriate to classify his actions in terms of the house, the garden, and the wood, and yet there seems no reason to suppose that he *individuates* these objects the way he does (it seems) his ball, sticks, the cats, and his water bowl. There is of course an act of individuation going on here: *we as theorists* studying the dog's behavior individuate the house, the wood, etc. Our theory can treat these as single entities. But there seems to be no reason to assume that Sam, the agent, has similar individuation capacities.

be referred to as *individuals*, denoted by $a, b, c, \ldots$ We shall take these *individuals* as given.

In addition to the individuation of individuals, a cognitive agent will be able to recognize that various *properties* hold or fail to hold of some or all of those individuals, and that various *relations* hold or fail to hold between some of those individuals. I shall use $P, Q, R, \ldots$ in order to denote the properties and relations the agent recognizes or discriminates. Just which properties and relations these are is determined by the agent or species of agent. This is not an issue that greatly affects the development of a calculus of information flow.

The notion of (digital) *information* we shall adopt in our study is:

$$\text{objects } a_1, \ldots, a_n \text{ do/do not stand in the relation } P.$$

Here P is some property that applies to n or more objects of certain kinds, and $a_1, \ldots, a_n$ are objects in the ontology that are appropriate for the respective argument places of the relation P. The identification of the objects $a_1, \ldots, a_n$ is *not* assumed to be part of the information. That is to say, information is taken to be in an *itemized*, conceptual form, and each item of information pertains to certain *given* objects. Since we place no restriction on the kind of property P might be, this is an extremely broad notion of information that encompasses pretty well any single item of *de re* information (i.e., information about one or more given individuals) that can be imagined.

Dretske's theory of information

The above observations suggest that in order to fully understand information, we have to examine in some detail the mechanism by which a signal can represent information. One framework for doing this, situation theory, forms the topic of the next lecture. However, in the 1980s, Dretske hit upon an interesting approach that comes half way between a situation theory type treatment of information — which analyzes the constraints that enable signals to represent information — and Shannon's theory — which says virtually nothing about the information encoded in a signal. What Dretske did was use some of Shannon's ideas to examine the relationship between signals and information assuming merely that *there exists* some systematic linkage between the two, but without analyzing that linkage. Here is Dretske's approach.

First, let me remark that Dretske restricted what we understand by information to items of the form that a single object a has property P for some a, P. I'll do the same in what follows.

According to Shannon, the amount of information generated by a state of affairs s is

$$I(s) = \log 1/p(s)$$

and the amount of information about s carried by a signal r is

$$I_s(r) = I(s) - E(r)$$

where $E(r)$ is the equivocation associated with the signal r. In Shannon's theory, these quantities have no real meaning on their own. Rather, they are intermediate terms formulated during the analysis that leads to Shannon's information measure. But, as Dretske observed, you can take these quantities and use them to form the basis of a qualitative theory of information content. Not by calculating actual numerical values; as we just observed, that is almost never possible. Rather, by concentrating instead on simply comparing those numbers: asking which of two is the greater. For instance, if s is the information that I live in California, then there's no meaningful way you can calculate $I(s)$. Likewise, if r is the information that I live in Lafayette, California, you cannot compute $I(r)$. But you will surely agree that $I(s) < I(r)$. (An important particular instance of this is when s is an information source and r is a signal emanating from s, when we ask whether $I_s(r) < I(s)$. If the answer is yes, then there is equivocation at the source and r does not carry all the information generated at s.)

Based on this observation, we may follow Dretske and develop a theory of *incremental* information. (Dretske did not use this term.)

It is worth mentioning that Dretske regarded the quantities $I(s)$ and $I_s(r)$ as being objective features of s and r, even though we have no way of determining their values.

We start with two fundamental observations that Dretske made.

> Basic Principle of Information Conveyance: It is by virtue of a signal s being F (for some property F) that s carries the information that s is P.

> Xerox principle: If A carries the information that B and B carries the information that C, then A carries the information that C.

The Xerox principle (this is Dretske's own name for this principle) is what underlies the fact that information *flows*, often through a variety of media. We shall use it to guide us as we formulate three regulatory principles concerning the flow of information. (These are not precisely articulated axioms, rather just guiding principles to help us formulate a definition of what it means to say a signal carries certain information.)

> Fidelity Principle: If a signal carries the information that a is P, then a must be P.

According to this principle, there is no such thing as "false information." Personally, I find this restriction unduly inhibitive; situation theory provides a mechanism to avoid it, while not losing its main thrust.

> Adequacy Principle: If a signal carries the information that a is P, then the signal carries as much information about a as is generated by a being P.

For the office example from the first lecture, the selection of one of the eight individuals to be their representative generates 3 bits of information. According to the Adequacy Principle, a signal carrying this information must have at least 3 bits. This seems obvious enough. But the Adequacy Principle clearly does not go far enough. Many signals with 3 or more bits will not carry the information about who the workers selected as their representative. Quantity alone is not enough. To carry that information, the signal must, so to speak, carry the right 3 bits. Dretske attempts to overcome this problem with the following principle:

> Relevancy Principle: If a signal carries the information that a is P, then the quantity of information the signal carries about a must include that quantity generated by a being P.

Dretske himself admits that this is somewhat vague. Nevertheless, as we shall see, he is nevertheless able to go on and formulate a definition of what it means to carry information about something. The situation-theoretic approach we'll see in the next lecture can be regarded as one way to make precise Dretske's Relevancy Principle.

> Nested Information Principle: If a signal carries the information that a is P, then the signal carries any information nested in the fact that a is P.

For example, if r carries the information that a is a square, then r carries the information that a is a rectangle, that it is a planar geometric figure, etc.

With these principles in mind, we can now formulate a definition of (incremental) information content:

> Let K be the information a receiver of a signal r already has about a source s. Then r carries the information that a is P if and only if $p(P(a) \mid r, K) = 1$ but $p(P(a) \mid K) < 1$.

First, notice that this definition of information content does indeed satisfy our four guiding principles. It also exhibits an intentionality that most would agree is a desirable property: If r carries the information that a is P, then just because P and Q are extensionally equivalent it does not follow that r carries the information that a is Q. For example, if P is the property of being the current President of the United States, then a message that carries the information that a certain individual a is P does not necessarily carry the information that a is Q, where Q is the property of being a linguistically impaired current President of the United States, even if, as a matter of fact, the individual a is linguistically impaired.

However, arguably the most significant requirement in Dretske's theory is the clause of the Relevancy Principle that specifies the identity of the information carried with that generated at the source. This is the one place where Dretske's account goes beyond solely quantitative considerations. Yet without it, his theory collapses. To my mind, this demonstrates the need for any theory of information to include an account of the mechanism by which a signal encodes information.

This, in turn, requires a formally defined notion of information distinct from any representation of information. The only domain I can see in which to locate that notion is mathematics. This is where situation theory begins.

References

[1] Devlin, K. *Logic and Information*, Cambridge University Press (1991)

[2] Dretske, F. *Knowledge and the Flow of Information*, Bradford Books, MIT Press (1981)

[3] Dretske, F. Private communication, Jan 29, 1988

ESSLLI 2001, Helsinki, Finland

The Mathematics of Information

Lecture 3: Introduction to Situation Theory

Keith Devlin*

Basic ontology of situation theory

In situation theory, information is always taken to be information *about* some situation, and is taken to be in the form of discrete items known as *infons*. Infons are of the form

$$\ll R, a_1, \ldots, a_n, 1 \gg \quad , \quad \ll R, a_1, \ldots, a_n, 0 \gg$$

where R is an n-place relation and $a_1, \ldots, a_n$ are objects appropriate for R.

The class of *compound infons* is constructed from the infons by closing under operations of conjunction and disjunction and bounded existential and universal quantification (over parameters).

Infons (or compound infons) are 'items of information'. They are not things that in themselves are true or false. Rather a particular item of information may be true or false about a situation.

Given a situation, s, and an infon σ, write

$$s \models \sigma$$

to indicate that the infon σ is 'made factual by' the situation s. Read this as s *supports* σ.

The claim $s \models \sigma$ is referred to as a *proposition*.

The objects (called *uniformities*) in the ontology include the following.

- *individuals*, denoted by $a, b, c, \ldots$
- *relations*, denoted by $P, Q, R, \ldots$
- spatial *locations*, denoted by $l, l', l'', l_0, l_1, l_2, \ldots$
- *temporal locations*, denoted by $t, t', t_0, \ldots$
- *situations*, denoted by $s, s', s'', s_0, \ldots$
- *types*, denoted by $S, T, U, V, \ldots$

*Center for the Study of Language and Information, Stanford University, Stanford, California 94305. devlin@csli.stanford.edu

- *parameters*, denoted by $\dot{a}$, $\dot{s}$, $\dot{t}$, $\dot{l}$, etc.

The agent-relative framework that 'picks out' the ontology is referred to as the *scheme of individuation* (appropriate for a study of that agent *vis á vis* information flow).

The intuition is that in our study of the activity (both physical and cognitive) of a particular agent or species of agent, we notice that there are certain regularities or *uniformities* that the agent either individuates or else discriminates in its behavior.

Relations are structured objects that have (in particular): *appropriateness conditions* and *minimality conditions*.

The *basic types*:

- TIM : the type of a temporal location
- LOC : the type of a spatial location
- IND : the type of an individual
- REL^n : the type of an n-place relation
- SIT : the type of a situation
- INF : the type of an infon
- TYP : the type of a type (see later)
- PAR : the type of an *parameter* (see later)
- POL : the type of a polarity (0 and 1)

For each basic type T other than PAR, there is an infinite collection $T_1, T_2, T_3, \ldots$ of *basic parameters*, used to denote arbitrary objects of type T.

The parameters T_i are sometimes referred to as T-*parameters*.

Notation: $\dot{l}$, $\dot{t}$, $\dot{a}$, $\dot{s}$, etc. to denote parameters (of type *LOC, TIM, IND, SIT*, etc.).

Given an object, x, and a type, T, we write

$$x : T$$

to indicate that the object x is *of* type T.

An *anchor* for a set, A, of basic parameters is a function defined on A, which assigns to each parameter T_n in A an object of type T.

If σ is a compound infon and f is an anchor for some of the parameters in σ, $\sigma[f]$ denotes the compound infon that results from replacing each parameter $\dot{a}$ in $\mathrm{dom}(f)$ by $f(a)$.

Restricted parameters are constructed as follows.

Let v be a parameter. A *condition* on v is a finite conjunction of infons. (At least one conjunct should involve v, otherwise the definition is degenerate.)

Given a parameter, v, and a condition, C, on v, define a new parameter, $v \upharpoonright C$, called a *restricted parameter*. $v \upharpoonright C$ will denote an object of the same type as v, that satisfies the requirements imposed by C (*in any situation where this applies*). (If C consists of a single parametric infon σ, we write $v \upharpoonright \sigma$ instead of $v \upharpoonright \{\sigma\}$.)

Let $r = v \upharpoonright C$ be a parameter. Given a situation s, a function f is said to be an *anchor* for r in s if:

1. f is an anchor for v and for every parameter that occurs free in C ;
2. for each infon σ in C: $s \models \sigma[f]$;
3. $f(r) = f(v)$.

There are two kinds of type-abstraction, leading to two kinds of types.

Situation-types. Given a *SIT*-parameter, $\dot{s}$, and a compound infon σ, there is a corresponding *situation-type*

$$[\dot{s} \mid \dot{s} \models \sigma],$$

the *type* of situation in which σ obtains. This process of obtaining a type from a parameter, $\dot{s}$, and a compound infon, σ, is known as *(situation-) type abstraction.* The parameter $\dot{s}$ is called the *abstraction parameter* used in this type abstraction.

For example,

$$[SIT_1 \mid SIT_1 \models \ll \text{running}, \dot{p}, LOC_1, TIM_1, 1 \gg]$$

Object-types. These include the basic types *TIM, LOC, IND,* REL^n, *SIT, INF, TYP, PAR,* and *POL*, as well as the more fine-grained uniformities described below.

Object-types are determined over some initial situation.

Let s be a given situation. If $\dot{x}$ is a parameter and σ is some compound infon (in general involving $\dot{x}$), then there is a type

$$[\dot{x} \mid s \models \sigma],$$

the *type* of all those objects x to which $\dot{x}$ may be anchored in the situation s, for which the conditions imposed by σ obtain.

This process of obtaining a type $[\dot{x} \mid s \models \sigma]$ from a parameter, $\dot{x}$, a situation, s, and a compound infon, σ, is called *(object-) type abstraction.* The parameter $\dot{x}$, is known as the *abstraction parameter* used in this type abstraction. The situation s is known as the *grounding* situation for the type.

In many instances, the grounding situation, s, is 'the world' or 'the environment' we live in (generally denoted by w). For example, the *type* of all people could be denoted by

$$[IND_1 \mid w \models \ll \text{person}, IND_1, \dot{l}_w, \dot{t}_{now}, 1 \gg]$$

Again, if s denotes Jon's environment (over a suitable time span), then

$$[\dot{e} \mid s \models \ll \text{sees, Jon}, \dot{e}, LOC_1, TIM_1, 1 \gg]$$

denotes the type of all those situations Jon sees (within s). This is a case of an object-type that is a type of situation. This example is not the same as a *situation-type.* Situation-types classify situations according to their internal structure, whereas in the type

$$[\dot{e} \mid s \models \ll \text{sees, Jon}, \dot{e}, LOC_1, TIM_1, 1 \gg]$$

the situation is typed from the outside.

Constraints may be natural laws, conventions, logical (i.e. analytic) rules, linguistic rules, empirical, law-like correspondences, etc. For example:

smoke means fire.

If S is the type of situations where there is smoke present, and S' is the type of situations where there is a fire, then an agent (eg. a person) can pick up the information that there is a fire by observing that there is smoke (a type S situation) and being aware of, or *attuned to*, the constraint that links the two types of situation.

This constraint is denoted by

$$S \Rightarrow S'$$

(This is read as "S *involves* S'.")

Another example: FIRE *means fire.*

This describes the constraint

$$S'' \Rightarrow S'$$

that links situations (of type S'') where someone yells the word FIRE to situations (of type S') where there is a fire.

Awareness of the constraint

FIRE *means fire*

involves knowing the meaning of the word FIRE and being familiar with the rules that govern the use of language.

The three types just introduced may be defined as follows:

$$\begin{aligned} S &= [\dot{s} \mid \dot{s} \models \ll \text{smokey}, \dot{t}, 1 \gg] \\ S' &= [\dot{s} \mid \dot{s} \models \ll \text{firey}, \dot{t}, 1 \gg] \\ S'' &= [\dot{u} \mid \dot{u} \models \ll \text{speaking}, \dot{a}, \dot{t}, 1 \gg \wedge \ll \text{utters}, \dot{a}, \textit{fire}, \dot{t}, 1 \gg] \end{aligned}$$

Notice that constraints links types, not situations.

On the other hand, any particular instance where a constraint is utilized to make an inference or modify behavior will involve specific situations (of the relevant types).

Thus constraints function by relating various regularities across actual situations.

Infon logic

We may form the *conjunction*, $\sigma \wedge \tau$, of two infons, σ, τ. The conjunction is not itself an infon, but a *compound infon*.

For any situation, s, we have

$$s \models \sigma \wedge \tau \text{ iff } s \models \sigma \text{ and } s \models \tau.$$

The *disjunction* of two infons, σ, τ, is a compound infon $\sigma \vee \tau$ such that for any situation s,

$$s \models \sigma \vee \tau \text{ iff } s \models \sigma \text{ or } s \models \tau \text{ (or both).}$$

We regard the above definitions as clauses in a recursive definition of compound infons.

If σ is an infon (or compound infon) that involves the parameter $\dot{x}$ and u is some set, then

$$(\exists \dot{x} \in u)\sigma$$

is a compound infon.

For any situation, s, that contains (as constituents) all members of u :

$s \models (\exists \dot{x} \in u)\sigma$ iff there is an anchor, f, of $\dot{x}$ to an element of u, such that $s \models \sigma[f]$.

The anchor, f, here may involve some *resource situation* other than s. f must assign to $\dot{x}$ an appropriate object in some anchoring situation, e, that supports the various infons that figure in the structure of $\dot{x}$.

Example. Let σ be the compound infon

$$\ll \text{tired}, \dot{c}, t_0, 1 \gg \wedge \ll \text{hungry}, \dot{c}, t_0, 1 \gg$$

where $\dot{c}$ is a parameter for a cat.

Let s be a room situation at time t_0 and u the set of individuals in s. Then:

$$s \models (\exists \dot{c} \in u)\sigma$$

iff there is an anchor, f, of $\dot{c}$ to some fixed object, c, in u (c necessarily a cat) such that $s \models \sigma[f]$, i.e. such that

$$s \models \ll \text{tired}, c, t_0, 1 \gg \wedge \ll \text{hungry}, c, t_0, 1 \gg$$

That is to say, $s \models (\exists \dot{c} \in u)\sigma$ iff there is a cat, c, in u that at time t_0 is tired and hungry in s.

The existence of the anchor, f, entails the existence of an associated *anchoring* (or *resource*) *situation*, e, such that (in particular)

$$e \models \ll \text{cat}, c, 1 \gg$$

In particular, c is a constituent of e.

Note that the object c has to be in the (room) situation, s, at time t_0 in order for the proposition

$$s \models \ll \text{tired}, c, t_0, 1 \gg \wedge \ll \text{hungry}, c, t_0, 1 \gg$$

to obtain.

If σ is an infon (or compound infon) that involves the parameter $\dot{x}$, and if u is some set, then

$$(\forall \dot{x} \in u)\sigma$$

is a compound infon.

For any situation, s, that contains (as constituents) all members of u:

$$s \models (\forall \dot{x} \in u)\sigma$$

iff, for all anchors, f, of $\dot{x}$ to an element of u, $s \models \sigma[f]$.

In the cases both of existential and universal quantification, the bounding set u may consist of all the objects of a certain kind that are in the situation s. Consequently, the definitions do provide a notion of 'unrestricted' quantification, but it is a notion of *situated* quantification. For example, when I truthfully assert

All citizens have equal rights

I am presumably quantifying over some country such as the United States, not the entire world, for which such a claim is not true.

Naive situation semantics

We analyze utterances in terms of three situations:

- Utterance situation
- Resource situation
- Focal situation

The *utterance situation*. This is the context in which the utterance is made and received.

If Jan says to Naomi

A man is at the door

the utterance situation, u, is the immediate context in which Jan utters these words and Naomi hears them.

The situation u includes both Jan and Naomi (for the duration of the utterance), and should be sufficiently rich to identify various salient factors about this utterance, such as the door that Jan is referring to.

This is probably the one in her immediate environment, but not necessarily. For instance, if Jan utters the sentence *A man is at the door* as part of a larger discourse, the situation u could provide an alternative door.

The connections between the utterance and the various objects referred to, are known as just that: *connections* (or *speaker's connections*). Thus

$$u \models \ll \text{utters, Jan, } \textit{A man is at the door}, l, t, 1 \gg \wedge \ll \text{refers-to, Jan, } \textit{the door}, D, l, t, 1 \gg$$

where D is a door that is fixed by u. The speaker's connections link the utterance (as part of u) of the phrase *the door* to the object D.

Resource situations. If Jan says

The man I saw running yesterday is at the door,

she is making use of a situation that she witnessed the day before, the one in which a certain man was running, in order to identify the man at the door. There is another situation, r, a situation that occurred the day before the utterance, and which Jan witnessed, such that

$$\begin{aligned} u \quad \models \quad & \ll \text{utters, Jan, } \Phi, l, t, 1 \gg \wedge \\ & \ll \text{refers-to, Jan, } \textit{the man}, M, l, t, 1 \gg \wedge \\ & \ll \text{refers-to, Jan, } \textit{the door}, D, l, t, 1 \gg \end{aligned}$$

where Φ is the sentence

The man I saw running yesterday is at the door

and where Jan is making use of r and the fact that M is the unique man such that (for some appropriate values of l', t')

$$r \models \ll \text{runs}, \ M, l', t', 1 \gg$$

Resource situations can become available for exploitation in various ways, such as:

1. by being perceived by the speaker;
2. by being the objects of some common knowledge about the world;
3. by being the way the world is;
4. by being built up by previous discourse.

The *focal situation*. That part of the world the utterance is about.

Features of the utterance situation serve to identify the focal situation. For instance, suppose Jan makes her utterance while peering out of the upstairs window at the house across the street. Then her utterance refers to the situation, s, that she sees, the situation at the house across the street, and we have

$$s \models \ll \text{present}, \ M, l, t, 1 \gg$$

where l is the location of the door and t is the time of the utterance.

Propositional content

By adopting an ontology that includes items of information (infons), we are able to capture the notion of *the information* encoded by a representation, and can account for the fact that the same information can be encoded by two quite different representations, using quite different representation schemas.

There are then three notions that are often treated as if they were somewhat interchangeable, but which situation theory regards as quite distinct (though related):

- information
- representations
- propositions.

In the case of a linguistic utterance, say Jon's utterance of the assertive sentence

Mary is running

the representation is the utterance itself, which we regard as a situation, call it u.

The *propositional content* of the utterance u is the proposition

$$e \models \sigma$$

where e is the focal situation, σ is the infon $\ll$ runs, $M, t_u, 1 \gg$, M denotes the individual Mary to whom Jon refers, t_u is the time of the utterance, and e is determined by various features of the utterance.

For example, e could be determined by Jon and the listener being part of some larger situation in which this individual Mary is running, or more generally by means of some other form of previously established context of utterance.

The propositional content is what might normally be referred to as the 'information conveyed by the utterance'.

Reference

Devlin, Keith, *Logic and Information*, Cambridge University Press (1991).

ESSLLI 2001, Helsinki, Finland

The Mathematics of Information

Lecture 4: Introduction to Channel Theory

Keith Devlin*

During the 1970s, Barwise and Seligman took the situation-theoretic approach to information and took away much of the technical machinery for handling situations, relations and types, to produce an abstract mathematical account of information flow.

Their motivating idea is that information flow is made possible by regularities in systems. Rather than develop machinery for analyzing those regularities, however, as situation theory does, instead they built a mathematical theory based on the mere existence of such regularities. The starting point is the notion of a classification.

Classifications

A *classification* is a structure $\boldsymbol{A} = \langle A, \Sigma_A, \models_A \rangle$, where A is a set of objects to be classified, called the *tokens* of $\boldsymbol{A}$, Σ_A is a set of objects used to classify the tokens, called the *types* of $\boldsymbol{A}$, and $\models_A$ is a binary relation between A and Σ_A which determines which tokens are classified by which types. We illustrate the classification relation as follows:

$$\begin{tikzcd} \Sigma_A \arrow[d, dash, "\models_A"] \\ A \end{tikzcd}$$

A familiar example to logicians is where the types are sentences of first-order logic and the tokens are mathematical structures, and $a \models \alpha$ is the relationship that the structure a is a model of the sentence α.

The first step is to develop machinery for discussing the “logic” of a system by means of which the system can support the flow of information.

Given a classification $\boldsymbol{A}$, a *sequent* is a pair (Γ, Δ) of sets of types of $\boldsymbol{A}$.

A token a of $\boldsymbol{A}$ is said to *satisfy* the sequent (Γ, Δ) if,

$$(\forall \alpha \in \Gamma)[a \models \alpha] \;\Rightarrow\; (\exists \alpha \in \Delta)[a \models \alpha]$$

*Center for the Study of Language and Information, Stanford University, Stanford, California 94305. devlin@csli.stanford.edu

We say that Γ *entails* Δ in $\boldsymbol{A}$, written $\Gamma \vdash_A \Delta$, if every token of $\boldsymbol{A}$ satisfies (Γ, Δ).

If $\Gamma \vdash_A \Delta$, then the pair (Γ, Δ) is said to be a *constraint* supported by the classification $\boldsymbol{A}$.

The set of all constraints supported by $\boldsymbol{A}$ is called the complete theory of $\boldsymbol{A}$, denoted by $\mathrm{Th}(\boldsymbol{A})$. The complete theory of $\boldsymbol{A}$ represents all the regularities supported by the system being modeled by $\boldsymbol{A}$.

Notice the following special cases of constraints:

- If α and β are both singletons, then $\alpha \vdash \beta$ is the claim that α logically entails β.
- The constraint $\vdash \alpha$, where the left-hand side is empty and α is a singleton, is the claim that α is necessarily true.
- The constraint $\alpha \vdash$, where the right-hand side is empty and α is a singleton, is the claim that no token is of type α.
- The constraint $\vdash \alpha, \beta$, where α, β is a doubleton, is the claim that every token is of (at least) one of the types α, β.
- The constraint $\alpha, \beta \vdash$, where the right-hand side is empty and the left-hand side is a doubleton, is the claim that the types α and β are mutually exclusive (i.e., no token is of type α and of type β).

Infomorphisms

In developing channel theory, Barwise and Seligman formulated three guiding principles concerning information flow:

- Information flow results from regularities in a distributed system.
- Information flow crucially involves both types and tokens.
- It is by virtue of regularities among connections that information about some components of a distributed system carries information about other components.

Let $\boldsymbol{A} = \langle A, \Sigma_A, \models_A \rangle$ and $\boldsymbol{C} = \langle C, \Sigma_C, \models_C \rangle$ be two classifications. An *infomorphism* between $\boldsymbol{A}$ and $\boldsymbol{C}$ is a pair $f = (f^\wedge, f^\vee)$ of functions that makes the following diagram commute:

$$\begin{tikzcd}
\Sigma_A \arrow[r, "f^\wedge"] \arrow[d, dash, "\models_A"] & \Sigma_C \arrow[d, dash, "\models_C"] \\
A & C \arrow[l, "f^\vee"]
\end{tikzcd}$$

This means that

$$f^\vee(c) \models_A \alpha \ \text{ iff } \ c \models_C f^\wedge(\alpha)$$

for all tokens c of $\boldsymbol{C}$ and all types α of $\boldsymbol{A}$. We refer to $f^\wedge$ as "f-up" and $f^\vee$ as "f-down". We take account of the fact that the functions $f^\wedge$ and $f^\vee$ act in opposite directions by writing

$$f : \boldsymbol{A} \rightleftarrows \boldsymbol{C}$$

For example, suppose we have two mathematical theories M_1 and M_2, with corresponding first-order languages L_1 and L_2, respectively, and suppose further it is possible to interpret M_1 within M_2. (For example, M_1 could be arithmetic and M_2 set theory.) That is to say, with each structure m_2 of M_2 we can associate a unique structure $m_1 = I^{\vee}(m_2)$ of M_1, and each sentence α_1 of L_1 has a unique canonical translation $\alpha_2 = I^{\wedge}(\alpha_1)$ of L_2. (In the case where M_1 is arithmetic and M_2 set theory, $I^{\vee}(m_2)$ is the arithmetic in the model m_2 of set theory and $I^{\wedge}(\alpha_1)$ is the standard translation of α_1 into the language of set theory using the von Neumann definition of the natural numbers and the usual set-theoretic definitions of the operations of arithmetic.) Then $I = (I^{\wedge}, I^{\vee})$ is an infomorphism and the following diagram commutes:

$$\begin{tikzcd}
L_1\text{-sentences} \arrow[r, "I^{\wedge}"] \arrow[d, dash, "\models_{L_1}"] & L_2\text{-sentences} \arrow[d, dash, "\models_{L_2}"] \\
L_1\text{-structures} & L_2\text{-structures} \arrow[l, "I^{\vee}"']
\end{tikzcd}$$

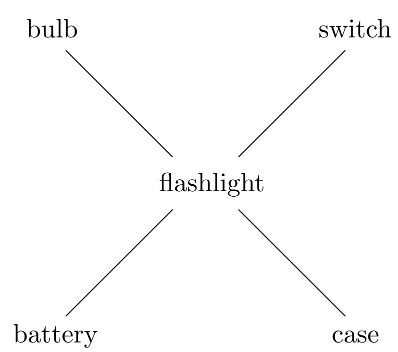
Based on the above example, given an infomorphism $f : \boldsymbol{A} \rightleftarrows \boldsymbol{C}$, we call $f^{\wedge}(\alpha)$ the *translation* of α and $f^{\vee}(c)$ the *particularization* of c.

Infomorphisms allow us to investigate the manner in which the parts of a system fit together to facilitate the flow of information. The key notion is that of an information channel.

An *information channel* consists of an indexed family $\mathcal{C} = \{f_i : \boldsymbol{A}_i \rightleftarrows \boldsymbol{C}\}_{i \in I}$ of infomorphisms with a common codomain $\boldsymbol{C}$, called the *core* of the channel.

The intuition is that the $\boldsymbol{A}_i$ are individual components of the larger system $\boldsymbol{C}$, and it is by virtue of being parts of the system $\boldsymbol{C}$ that the constituents $\boldsymbol{A}_i$ can carry information about one another.

For example, consider a simple flashlight comprising a case, a bulb, a switch, and a battery. We can illustrate this system as follows:



The structure of the flashlight means that various components can carry information about others. For example, if the switch is of type ON and the battery is of type CHARGED then the bulb will be of type LIT.

In terms of an information channel, the flashlight can be represented like this:

$$
\begin{tikzcd}
\boldsymbol{A}_1 \arrow[dr, "f_1"'] & & \boldsymbol{A}_2 \arrow[dl, "f_2"] \\
& \boldsymbol{C} & \\
\boldsymbol{A}_4 \arrow[ur, "f_4"] & & \boldsymbol{A}_3 \arrow[ul, "f_3"']
\end{tikzcd}
$$

Information flow

Based on the three basic principles they formulated (and other considerations), Barwise and Seligman proposed the following definition:

> Suppose $\boldsymbol{A}$ and $\boldsymbol{B}$ are constituent classifications in an information channel with core $\boldsymbol{C}$. A token a being of type α in $\boldsymbol{A}$ carries the information that a token b is of type β in $\boldsymbol{B}$ relative to the channel $\boldsymbol{C}$ if a and b are connected in $\boldsymbol{C}$ and the translation of α entails the translation of β in $\mathrm{Th}(\boldsymbol{C})$.

For example, in the case of the flashlight, suppose the classifications for the bulb, switch, battery, and case are everyday, common sense ones (bulbs have types LIT, UNLIT, and BROKEN, switches have types ON and OFF, etc.) and the classification of the flashlight is technical, involving principles of engineering, the laws of physics, etc. What does it mean to say that the switch of a particular flashlight being on carries the information that the bulb is lit?

Well, using the notation of the above diagram, the $\boldsymbol{A}_2$-type (i.e., the everyday switch-type) ON has a translation $f_2^\wedge(\textsc{on})$ in $\boldsymbol{C}$. Likewise, the $\boldsymbol{A}_1$-type (i.e., the everyday bulb-type) LIT has a translation $f_1^\wedge(\textsc{lit})$ in $\boldsymbol{C}$. The switch being on carries the information that the bulb is lit by virtue of the inference

$$f_2^\wedge(\textsc{on}) \vdash_{\boldsymbol{C}} f_1^\wedge(\textsc{lit})$$

Notice that the types in $\boldsymbol{C}$ provide the logical structure — the regularities — that gives rise to information flow, but information only flows in the context of a particular token c of $\boldsymbol{C}$ (i.e., a particular flashlight), for this is what provides the specific connections required to facilitate information flow. Specifically, if switch s is connected to bulb b by flashlight c, then s being on carries the information that b is lit.

We may further elucidate the way channels work by showing that the above definition of information flow satisfies Dretske's principle of veridicality: if $a \models_{\boldsymbol{A}} \alpha$ carries the information that $b : \beta$, then $b \models_{\boldsymbol{B}} \beta$. To see this, let c be the token in $\boldsymbol{C}$ that connects a and b. (The down-components of the relative infomorphisms map c to a and b.) If α' is the translation of α and β' that of β (i.e., the images of α and β under the up-components of the infomorphisms), then c must satisfy $\alpha' \vdash_{\boldsymbol{C}} \beta'$. Now, since c is the image of a under an infomorphism, $c \models_{\boldsymbol{C}} \alpha'$. Hence, applying the inference in $\boldsymbol{C}$, $c \models_{\boldsymbol{C}} \beta'$. Thus, applying the infomorphism that takes us from $\boldsymbol{C}$ to $\boldsymbol{B}$, $b \models_{\boldsymbol{B}} \beta$, as claimed.

To show that the Barwise-Seligman definition satisfies Dretske's Xerox Principle, we need a method for combining classifications.

Given classifications $\boldsymbol{A}$ and $\boldsymbol{B}$, we define the *colimit* $\boldsymbol{A}+\boldsymbol{B}$ as follows. The tokens of $\boldsymbol{A}+\boldsymbol{B}$ consist of pairs (a, b) of tokens from each. The types of $\boldsymbol{A}+\boldsymbol{B}$ consists of of the types of both, except that, if there are any types in common, we make two distinct (indexed) copies in order not to confuse them.

There are natural infomorphisms $\sigma_A : \boldsymbol{A} \rightleftarrows [\boldsymbol{A}+\boldsymbol{B}]$ and $\sigma_B : \boldsymbol{B} \rightleftarrows [\boldsymbol{A}+\boldsymbol{B}]$ defined thus:

1. $\sigma_A^\wedge(\alpha) = \alpha_A$ (the $\boldsymbol{A}$-copy of α), for each type α of $\boldsymbol{A}$.
2. $\sigma_B^\wedge(\beta) = \beta_B$, for each type β of $\boldsymbol{B}$.
3. for each token (a, b) of $\boldsymbol{A}+\boldsymbol{B}$, $\sigma_A^\vee((a, b)) = a$ and $\sigma_B^\vee((a, b)) = b$.

The name "colimit" comes from category theory. The classification $\boldsymbol{A}+\boldsymbol{B}$ has the property that, given any classification $\boldsymbol{C}$ and infomorphisms $f : \boldsymbol{A} \rightleftarrows \boldsymbol{C}$, $g : \boldsymbol{B} \rightleftarrows \boldsymbol{C}$, there is a unique infomorphism $h = f + g$ such that the following diagram commutes:

$$\begin{tikzcd}
& \boldsymbol{C} & \\
\boldsymbol{A} \arrow[ur, "f"] \arrow[r, "\sigma_A"'] & \boldsymbol{A}+\boldsymbol{B} \arrow[u, "h"] & \boldsymbol{B} \arrow[ul, "g"'] \arrow[l, "\sigma_B"]
\end{tikzcd}$$

[Remember that an infomorphism consists of a pair of functions going in opposite directions. Commutativity in this case applies to both directions.]

The definition of h is obvious: on tokens, $h^\vee(c) = (f^\vee(c), g^\vee(c))$; on types α_A, $h^\wedge(\alpha_A) = f^\wedge(\alpha)$; on types α_B, $h^\wedge(\alpha_B) = g^\wedge(\alpha)$.

To verify the Xerox Principle, now, if $a \models_{\boldsymbol{A}} \alpha$ carries the information that $b : \beta$ in $\boldsymbol{B}$, it does so by some channel with core $\boldsymbol{C}_1$. If $b \models_{\boldsymbol{B}} \beta$ carries the information that $d : \delta$ in $\boldsymbol{D}$, it does so by some channel with core $\boldsymbol{C}_2$. Let $\boldsymbol{C} = \boldsymbol{C}_1 + \boldsymbol{C}_2$. Then it is not hard to see that $a \models_{\boldsymbol{A}} \alpha$ carries the information that $d : \delta$ in $\boldsymbol{D}$ by way of the channel $\boldsymbol{C}$.

Reasoning at a distance

One obvious drawback with the approach so far is that is assumes we have complete information about $\mathrm{Th}(\boldsymbol{C})$, the theory of the channel core. This is fine for a theoretical analysis, but in practice this is generally not the case. Often, we use commonsense knowledge of the core in order to attribute information about one component to another. We now analyze such reasoning at a distance.

By way of illustration, consider a situation where an agent operating in a local environment uses local information in order to obtain information about and manipulate a distant environment. For example, an engineer in the control room of a nuclear power plant uses various monitors and gauges to obtain information about the reactor and uses switches and dials to control the reactor. In this case there are three classifications: the proximal classification $\boldsymbol{P}$, the distal classification $\boldsymbol{D}$, and the channel classification $\boldsymbol{C}$ that connects the two. The situation is illustrated below:

$$\begin{tikzcd} & \boldsymbol{C} & \\ \boldsymbol{P} \arrow[ur, "p"] & & \boldsymbol{D} \arrow[ul, "d"'] \end{tikzcd}$$

In using information in $\boldsymbol{P}$ to reason about $\boldsymbol{D}$, the operator makes (implicit) use of the infomorphisms $p : \boldsymbol{P} \rightleftarrows \boldsymbol{C}$ and $d : \boldsymbol{D} \rightleftarrows \boldsymbol{C}$. When we trace the information flow from a sensor in $\boldsymbol{D}$ to a gauge in $\boldsymbol{P}$, we use the infomorphism p from $\boldsymbol{P}$ to $\boldsymbol{C}$ followed by the infomorphism d backwards from $\boldsymbol{C}$ to $\boldsymbol{D}$.

We can develop the machinery we need to analyze this situation by considering a general situation of an infomorphism $f : \boldsymbol{A} \rightleftarrows \boldsymbol{B}$. Imagine someone who has to reason about tokens on one side using the natural theory of the other. We want to see how constraints (inferences) in one classification give rise to constraints in the other. That is, we need to formulate rules that tell us how to an inference in one classification corresponds to an inference in the other.

A good example to keep in mind is where $\boldsymbol{A}$ is Peano Arithmetic (PA) and $\boldsymbol{B}$ is set theory (say ZFC).

If Γ is a set of types of $\boldsymbol{A}$, we denote by Γ^f the set of translations of types in Γ.

If Γ is a set of types of $\boldsymbol{B}$, we denote by Γ^{-f} the set of types of $\boldsymbol{A}$ whose translations are in Γ.

The following two inference rules allow us to pass from one classification to another:

$$f\text{-}\textbf{Intro}: \quad \frac{\Gamma^{-f} \vdash_A \Delta^{-f}}{\Gamma \vdash_B \Delta}$$

$$f\text{-}\textbf{Elim}: \quad \frac{\Gamma^{f} \vdash_B \Delta^{f}}{\Gamma \vdash_A \Delta}$$

The first rule allows us to go from a sequent of $\boldsymbol{A}$ to a sequent of $\boldsymbol{B}$; the second rule allows us to go the other way round. In a moment we'll examine the validity of these rules, but first let's use the number theory and set theory example to get a sense of what they say.

In the case of number theory and set theory, f-**Intro** says that, if we take a valid sequent in PA, its translation into set theory is a valid sequent in ZFC. f-**Elim** says that if we take a sequent of set theory that happens to be a translation of a sequent in number theory, and if the sequent is valid in ZFC, then the original sequent is valid in PA.

Now what can we say about the validity of these two rues: what do they preserve?

f-**Intro** preserves validity. For if c were a counterexample to $\Gamma \vdash_B \Delta$, $f(c)$ would be a counterexample to $\Gamma^{-f} \vdash_A \Delta^{-f}$. This is obvious for the case of PA and ZFC.

f-**Elim** does not preserve validity. There can be a valid constraint $\Gamma^f \vdash_B \Delta^f$ such that $\Gamma \vdash_A \Delta$ has a counterexample, although no counterexample can be of the form $f(c)$ for a token c of $\boldsymbol{B}$. For example, theorems of set theory in the language of number theory are reliable as long as we are only interested in models of number theory that are parts of models of set theory, but for other models the theorems are unreliable.

f-**Intro** does not preserve nonvalidity. For example, there are nontheorems of PA whose translations are theorems of ZFC; the consistency statement for PA is one such.

f-**Elim** does preserve nonvalidity. For instance, if a translation into set theory of a statement of number theory is false in ZFC, then the original statement must be false in PA.

Turning to systems now, the validity preserving nature of the f-**Intro** rule tells us that any constraint that holds for a component of a system translates to a constraint that holds for the system. Using f-**Elim**, however, we see that any constraint about the whole system gives a constraint about the components but only guarantees that it holds on those tokens that really are the components of some token of the whole system.

In the case of the proximal–distal channel

$$\begin{tikzcd} & \boldsymbol{C} & \\ \boldsymbol{P} \arrow[ur, "p"] & & \boldsymbol{D} \arrow[ul, "d"'] \end{tikzcd}$$

we considered earlier, we can now examine what happens when we use the complete theory of the proximal classification $\boldsymbol{P}$ to reason about the distal classification $\boldsymbol{D}$.

The first step is to go from $\boldsymbol{P}$ to $\boldsymbol{C}$. Since p-**Intro** preserves validities, the translated theory we obtain on $\boldsymbol{C}$ is sound. But it may not be complete; there may be constraints of $\boldsymbol{C}$ that we miss.

Going from $\boldsymbol{C}$ to $\boldsymbol{D}$ now, using d-**Elim** means that we lose soundness (in addition to the completeness we lost in the first stage). A sequent about distal tokens obtained from a constraint about proximal tokens in this way is guaranteed to apply to distal tokens that are connected to a proximal token in the channel, but there are no guarantees about other distal tokens.

We shall track what is going on using the notion of a local logic. This generalizes the notion of the complete theory of a classification.

Local logics

A *local logic* $\mathcal{L} = \langle \boldsymbol{A}, \vdash_{\mathcal{L}}, N_{\mathcal{L}} \rangle$ consists of a classification $\boldsymbol{A}$, a set $\vdash_{\mathcal{L}}$ of sequents (satisfying certain structural rules) involving the types of $\boldsymbol{A}$, called the *constraints* of $\mathcal{L}$, and a subset $N_{\mathcal{L}} \subseteq \boldsymbol{A}$, called the *normal tokens* of $\mathcal{L}$, which satisfy all the constraints of $\vdash_{\mathcal{L}}$.

A local logic $\mathcal{L}$ is *sound* if every token is normal; it is *complete* if every sequent that holds of all normal tokens is in the consequence relation $\vdash_{\mathcal{L}}$.

A sound and complete local logic is essentially a classification. Using infomorphisms, however, we can move local logics around from one classification to another in a way that does not preserve soundness and completeness.

Given an infomorphism $f : \boldsymbol{A} \rightleftarrows \boldsymbol{B}$ and a logic $\mathcal{L}$ on one of these classifications, we obtain a natural logic on the other. If $\mathcal{L}$ is a logic on $\boldsymbol{A}$, then $f[\mathcal{L}]$ denotes the logic on $\boldsymbol{B}$ obtained from $\mathcal{L}$ by f-**Intro**. If $\mathcal{L}$ is a logic on $\boldsymbol{B}$, then $f^{-1}[\mathcal{L}]$ denotes the logic on $\boldsymbol{A}$ obtained from $\mathcal{L}$ by f-**Elim**.

For any binary channel $\boldsymbol{C}$ as above, we define the local logic $\mathrm{Log}_{\boldsymbol{C}}(\boldsymbol{D})$ on $\boldsymbol{D}$ induced by that channel as

$$\mathrm{Log}_{\boldsymbol{C}}(\boldsymbol{D}) = d^{-1}[p[\mathrm{Log}(\boldsymbol{P})]]$$

where $\mathrm{Log}(\boldsymbol{P})$ is the sound and complete logic of the proximal classification $\boldsymbol{P}$. This logic builds in the logic implicit in the complete theory of the classification $\boldsymbol{P}$

As we have observed, it may be that $\mathrm{Log}_{\boldsymbol{C}}(\boldsymbol{D})$ is neither sound nor complete. But it is what is available in order to reason about $\boldsymbol{D}$ in $\boldsymbol{P}$.

It can be proved that every local logic on a classification $\boldsymbol{D}$ is of the form $\mathrm{Log}_{\boldsymbol{C}}(\boldsymbol{D})$ for some binary channel $\boldsymbol{C}$.

How information really flows

With the notion of a local logic available, let's now look again at Barwise and Seligman's definition of how information flows:

> Suppose $\boldsymbol{A}$ and $\boldsymbol{B}$ are constituent classifications in an information channel with core $\boldsymbol{C}$. A token a being of type α in $\boldsymbol{A}$ carries the information that a token b is of type β in $\boldsymbol{B}$ relative to the channel $\boldsymbol{C}$ if a and b are connected in $\boldsymbol{C}$ and the translation of α entails the translation of β in $\mathrm{Th}(\boldsymbol{C})$.

In many cases, $\boldsymbol{A}$ and $\boldsymbol{B}$ will consist of everyday folk theories (say of bulbs and switches in flashlights) whereas $\boldsymbol{C}$ will be a scientific/engineering theory — everything it takes to explain why flashlights work. Now, in reasoning about why the switch being moved to ON will carry the information that the light is LIT (or vice versa), a typical flashlight user (as opposed to a scientist or engineer) will use the folk theories. According to the above definition, however, the logic flow depends not on the folk theories themselves but on their translations into the scientific/engineering theory $\boldsymbol{C}$. Although this is arguably correct at a theoretical level — it is after all the science and the engineering that tells us exactly how flashlights work — this does not really capture the reasoning of the user. This we can do using local logics.

First, we note that the folk theory of how flashlights work is an amalgam of the folk theories of switches, bulbs, batteries, cases, and whatever. We represent this in our theory by using a natural extension of the colimit construction. Given any channel $\boldsymbol{C} = \{f_i : \boldsymbol{A}_i \rightleftarrows \boldsymbol{C}\}_{i\in I}$, we can represent it by a single infomorphism $f : \boldsymbol{A} \rightleftarrows C$ by taking the sum $\boldsymbol{A} = \sum_{i\in I} \boldsymbol{A}_i$ and the sum $f = \sum_{i\in I} f_i$ of the f_i. Given any logic $\mathcal{L}$ of the core, we can use f-**Elim** to obtain a local logic $f^{-1}[\mathcal{L}]$ on $\boldsymbol{A}$. It is this logic, with its constraints and normal tokens, that captures the information flow in the channel from the flashlight user's perspective. Or, as Barwise and Seligman put it [p.41], the local logic $f^{-1}[\mathcal{L}]$ is the "what" of information flow, the channel is the "why."

Let's look at the flashlight example in a bit more detail. We'll restrict our attention to the switch and the bulb for simplicity. Let $f : \boldsymbol{B} \rightleftarrows \boldsymbol{F}$ represent the part-whole relation between bulbs ($\boldsymbol{B}$) classified in commonsense ways and flashlights ($\boldsymbol{F}$) classified scientifically. Likewise, let $g : \boldsymbol{S} \rightleftarrows \boldsymbol{F}$ represent the part-whole relation between switches ($\boldsymbol{S}$) classified in commonsense ways and flashlights ($\boldsymbol{F}$) classified scientifically.

Putting these two infomorphisms together we get an infomorphism $h = f + g$ from $\boldsymbol{B} + \boldsymbol{S}$ to $\boldsymbol{F}$. Given a flashlight token x, $h(x) = (f(x), g(x))$, where $f(x)$ and $g(x)$ are the bulb and switch of x. Given a type Γ of $\boldsymbol{F}$, $h(\Gamma)$ is the disjoint union of $f(\Gamma)$ and $g(\Gamma)$.

Suppose now that $\boldsymbol{B}$ supports the constraint

$$\text{LIT} \vdash_{\boldsymbol{B}} \text{LIVE}$$

It is easily seen that this is a constraint of $\boldsymbol{B} + \boldsymbol{S}$. Then, whatever the classification $\boldsymbol{F}$ of flashlights is, and whatever h does to these types, we have

$$h(\text{LIT}) \vdash_{\boldsymbol{F}} h(\text{LIVE})$$

This is because h-**Intro** preserves validity.

Now let's go the other way round. Suppose the classification $\boldsymbol{F}$ supports the constraint

$$\text{ILLUM} \vdash_{\boldsymbol{F}} \text{ELEC}$$

where ILLUM is the technical property of an electrical component emitting photons and ELEC is the technical property of a component carrying electric current. Thus ILLUM $= h(\text{LIT})$ and ELEC $= h(\text{ON})$ (let us assume). Applying h-**Elim**, we get

$$\text{LIT} \vdash_{\boldsymbol{B}+\boldsymbol{S}} \text{ON}$$

However, this sequent is not valid. There are pairs (b, s) of switches and bulbs such that s is on but b is not lit. Indeed, there are many such pairs. The above inference only holds (qua an inference) for pairs that are tokens of the same flashlight. These pairs are the normal tokens of the logic obtained by h-**Elim**. This exemplifies our earlier observation that, in general, h-**Elim** does not preserve validity.

Handling exceptions

In examining the flashlight example, we assumed that all components were in working order. Thus, we did not question the validity of the constraint "If the switch is on, then the bulb is lit." But, of course, in real life this is not necessarily true. For instance, this constraint fails if the battery is dead. This is an example where the weakening rule

$$\frac{\alpha \vdash \gamma}{\alpha, \beta \vdash \gamma}$$

fails. (It is generally valid for mathematical reasoning.) The validity of the constraint

The switch being on entails the bulb is lit

does not imply the validity of the constraint

The switch being on and the battery being dead entails the bulb is lit

The channel-theoretic explanation for this phenomenon is that the introduction of an exception such as the battery being dead changes the channel to one where the relevant classifications have more types. Within a given channel, the Rule of Weakening does in fact hold.

For example, suppose we analyze the flashlight in terms of channels $\mathcal{F}$ and $\mathcal{F}'$ having core classifications $\boldsymbol{F}$ and $\boldsymbol{F}'$, respectively. Suppose that $\mathcal{F}'$ and $\mathcal{F}$ are identical apart from the fact that $\mathcal{F}'$ contains tokens where there are dead batteries but $\mathcal{F}$ does not. Then $\mathcal{F}'$ is a *refinement* of $\mathcal{F}$, which means that there is an infomorphism $r : \boldsymbol{F}' \rightleftarrows \boldsymbol{F}$ such that the following diagram commutes

$$
\begin{tikzcd}
& \boldsymbol{F} & \\
& \boldsymbol{F}' \arrow[u, "r"] & \\
\boldsymbol{B} \arrow[uur, "f"] \arrow[ur, "f'"] & & \boldsymbol{S} \arrow[uul, "g"'] \arrow[ul, "g'"']
\end{tikzcd}
$$

(Take r to be the identity on both types of $\mathcal{F}'$ and tokens of $\mathcal{F}$.)

Using r-**Intro**, any constraint of $\boldsymbol{F}'$ will yield a constraint of $\boldsymbol{F}$; i.e., any sequent that holds in $\mathcal{F}'$ will hold in $\mathcal{F}$. However, since r-**Elim** does not preserve validity, there may be constraint of $\mathcal{F}$ that are not valid sequents in $\mathcal{F}'$. This will in fact be the case if there is a flashlight x that has a dead battery. Relative to the channel $\mathcal{F}$, a flashlight switch being closed does carry the information that the bulb is lit; relative to the channel $\mathcal{F}'$, however, this is not the case. In symbols

$$h(\textsc{on}) \vdash_{\boldsymbol{F}} h(\textsc{lit})$$

but

$$h'(\textsc{on}) \not\vdash_{\boldsymbol{F}'} h'(\textsc{lit})$$

(where $h' = f' + g'$).

What channel theory does not do

Channel theory does an excellent job in achieving its aim: to provide a mathematical model of information flow that captures the way agents reason about the world using partial information. However, it was not developed as a tool to be used directly in real world reasoning. In that respect, it is very much like Dretske's theory that preceded it. In contrast, Shannon's theory and situation theory were both developed to be used — by communications engineers in the case of Shannon's theory and by social and computing scientists in the case of situation theory. By eliminating the machinery for handling types and constraints, channel theory was able to develop as an elegant mathematical theory. But it is, of course, precisely the types and constraints apparatus that are required for analyzing actual instances of information flow in the real world. Thus, situation theory and channel theory provide an excellent complementary pair of linked ways to approach information flow.

If this were a mathematics meeting, in my final lecture I would go into the mathematics of channel theory, proving various theorems about the theory. For this audience, however, I shall devote the final lecture to some applications of situation theory.

Reference

Barwise, Jon and Seligman, Jerry, *Information Flow: The Logic of Distributed Systems*, Cambridge University Press (1997).

ESSLLI 2001, Helsinki, Finland

The Mathematics of Information

Lecture 5: Applications of Situation Theory

Keith Devlin*

1 The Liar Paradox

The Liar Paradox arises when a person stands up and says, "This assertion is false." Is the speaker telling the truth or telling a lie? At first blush, it seems that the situation is paradoxical: if he is telling the truth, then he is not; if he is not telling the truth, then he is. In fact, that first blush lasted for almost two thousand years. First formulated by the ancient Greeks, the puzzle was not resolved until 1986, when Jon Barwise and John Etchemendy found an elegant solution using situation theory. As in the case of other famous paradoxes — such as Russell's paradox in set theory — the heart of the problem was an unacknowledged context. And just as the solution to Russell's paradox led to a better understanding of sets and to the formulation of suitable axioms of set theory, so too the solution to the Liar Paradox has lessons for us in terms of our understanding of language and communication.

To begin the analysis, suppose person a stands up and says, "This assertion is false." Let L denote the sentence uttered and let p be the proposition a makes by uttering L. Then a's utterance of the phrase "This assertion" refers to the claim p. It follows that, in uttering the sentence "This assertion is false," a is making the claim 'p is false'. Hence p and 'p is false' are one and the same:

(1) $$p = [p \text{ is false}]$$

In making the claim p — which is about that very claim — a must be making implicit reference to the context in which the assertion is made. Let c denote that context. Thus, a's utterance of the phrase "This assertion" refers to the claim $c \models p$. In other words, p must *be* the same as $c \models p$, since both are what a refers to by uttering the phrase "This assertion." Thus:

(2) $$p = [c \models p]$$

Suppose that a's assertion is true. In other words, p is true. Using formula (2), we can express this as

(3) $$c \models p$$

By formula (1), we can replace p in formula (3) by [p is false] to obtain:

(4) $$c \models [p \text{ is false}].$$

*Center for the Study of Language and Information, Stanford University, Stanford, California 94305. devlin@csli.stanford.edu

This is a contradiction: formula (3) tells us that p is true in the context c and formula (4) tells us that p is false in the same context c. Hence a's claim cannot be true.

Thus a's claim is false, i.e., p is false.

But what is the context for this last statement? Nothing we know provides a context for the statement 'p is false'. An obvious guess would be that c is itself the appropriate context. If so, then

$$(*) \qquad c \models [p \text{ is false}]$$

By formula (1), this can be rewritten as

$$(**) \qquad c \models p$$

But now we are in the same contradictory situation as we were in the previous case. Conclusion: c cannot be the appropriate context.

We do not have a paradox here, however, rather a theorem: (1) The person who utters the Liar sentence is uttering a falsehood, and (2) the context in which the claim of falsehood is made cannot be the same as the context in which the Liar sentence was uttered.

2 A mathematical model of manufacturing processes

[The work described in this section was part of a research project Keith Devlin carried out in collaboration with KBSI, Inc., with the support of the Advanced Research Projects Agency under Agile Manufacturing Pilot Program SOL BAA94-31A.]

In their paper [6], Menzel and Mayer describe a process (a manufacturing process, a business process, a social process, etc.) as "a structured collection of activities." As they point out, this is not so much a definition as a declaration of a starting point for an investigation. Indeed, there is currently no universally accepted definition of what a process is, though of course we all know one when we see one. One of the goals of this paper is to provide a formal, mathematical definition of a process.

In the absence of a formal definition — but in the presence of an established intuitive understanding — I shall illustrate my development by means of a simple example of a manufacturing process, a slight variant of the one used by Menzel and Mayer in their earlier paper [5]: the Paint–Dry Process (PD), specified by means of a flow chart in Figure 1. In this process, an item is subjected to three activities: painting, drying, and examining, interspersed by being placed in a queue. The examination results in the item either being released from the process or subjected to the same three activities once again. (A more realistic example would probably subject an item rejected by the examination to some sort of cleaning or scrubbing prior to repainting, but this does not affect my development, so I'll keep things simple.)

Though processes come in many shapes and sizes, this simple example is in many ways paradigmatic, illustrating the main problems that one faces in trying to formulate a formal definition of a process. Accordingly, it will be useful to begin this account with an examination of those problems, as they arise in the PD example.

If there was no examination stage and each part simply went through the process cycle once, a formal account would present no difficulty. Indeed, the flow chart would itself be a faithful, formal representation of the process. Likewise, if each item went through the queue–paint–dry cycle a fixed number N of times, a formal representation would seem to be relatively unproblematic. The case $N = 2$ (let's call this process PD2) is illustrated in Figure 2. Each of the three dashed-line boxes indicate that there is a single activity (queuing, painting, and

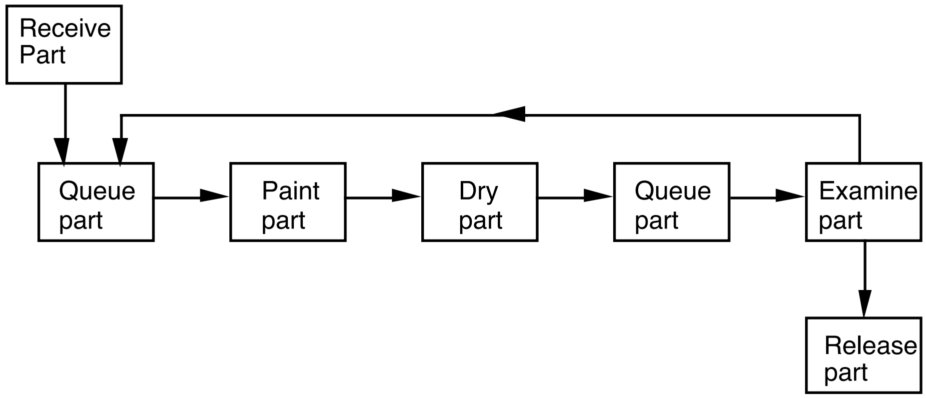


Figure 1: The Process PD

drying) taking place (probably at a single location, using the same apparatus), with each item being acted upon twice in succession.

And with that diagrammatic device, we already meet one of the fundamental questions that have to be resolved in order to develop a mathematical model of a process: What kind of entity is an 'activity'? When we write the term 'Paint part' in a box, are we representing an actual event, the painting of a single item at a particular time, or do we mean a *type* of event, the type of event that consists of an item (possibly of a single type) being painted at some time (possibly in the same way for each item, on each occasion)?

Most people who have looked at processes opt for the latter, that (at least in the context of processes) an activity is a *type*. I intend to adopt the same approach. In which case, the representation of PD2 in Figure 2 is misleading. The two boxes labeled 'Paint part', for instance, are enclosed in a single dashed-line box. The dashed-line box indicates that the two boxes enclosed denote two events of the same activity type. So the two 'Paint part' boxes appear to denote actual events, with the dashed-line box that encloses them denoting their common activity type. And yet, in the representation of PD in Figure 1, all the boxes clearly denote activity types, which seems right.

Of course, we could dig ourselves out of this particular hole by agreeing that all the boxes in

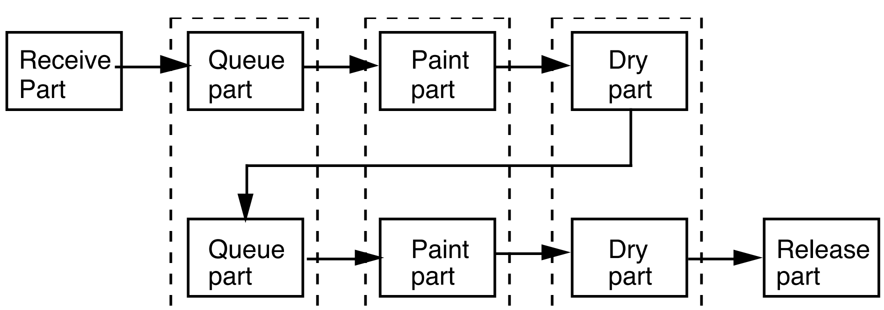


Figure 2: The Process PD2

Figure 2 denote activity types, and that enclosing two boxes within a dashed-line box indicates that the two types enclosed are really the same type. But let's leave the issue of how best to draw diagrams and move on to the question of finding a mathematical model of processes.

One feature of processes that is self-evident both from our intuitive understanding of processes and from the representational notations just discussed, is that it is almost entirely a matter of point-of-view as to what constitutes a 'process' and what an 'activity'. In the PD example, we regarded the painting, queuing, drying, and examining as activities and the entire sequence of those activities as a process. But we could equally regard PD as a single activity, that of 'finishing' the item, an activity that is part of an entire manufacturing process (sic). Similarly, we could of course probably regard each of 'Paint part', 'Dry part', and 'Examine part' as a process in its own right.

Thus, whatever framework we use to model processes, both processes and their constituent activities should be modeled by the same kind of mathematical objects. Moreover, our model needs to distinguish between the activities and processes, which are *types*, and their particular instantiations (or *activations*) when the activity or process is applied (to a particular object or objects) on a particular occasion.

With those preliminary observations out of the way, the fundamental question to be answered is what kind of event-type constitutes an *activity* (-type)? What distinguishes those event-types we can refer to as 'activities' from other kinds of event-types?

If we are to allow for the many different kinds of 'process' that can arise, then the only answer, as far as I can see, is that an activity-instance has a start-time and a finish-time, and durates between those two times. I shall take this as my starting point, and constuct a model of processes—more precisely, I construct a framework that can be used to model any process—with situation theory as the underlying mathematics. Once I have stated the main definitions, I shall illustrate them by means of our PD example.

By an *activity instance* we mean a situation s that:

1. has a *start time* t_s^b (b for 'begin');
2. has a *finish time* t_s^f;
3. has continuous temporal duration over the time-interval $[t_s^b, t_s^f]$.

An *activity* (or, more fully, an *activity type*) is a situation type such that any situation of that type is an activity instance. In general, we do not require that all situations of the type have the same start-time and finish time, though for some activities this may be the case.

A *process* is just another way to refer to an activity. The actual choice of terminology depends on the context of the discussion. In particular, a suitably structured sequence of activities is referred to as a process, and the constituent activities of a process are referred to as activities. A connected subsequence of the activities in a process is sometimes referred to as a *subprocess* of the given process.

'Pauses' are particular examples of activities. Given times t_0 and t_1, with $t_0 < t_1$, the *pause* $P = P(t_0, t_1)$ is the type of those activity instances whose start time is t_0, whose finish time is t_1, and which are otherwise unrestricted. [In order for a suitably structured sequence of activities to be a process (i.e. to be itself an activity), the start and finish times of the activities in the sequence must match up; more precisely, the start-time of each non-initial activity in the process must coincide with the finish-time of the preceding activity. (Recall that an activity must durate from its start-time to its finish-time.) To achieve this in our model, we take pauses to be legitimate activities.]

In order to capture within our model the way that a process breaks up into a structured sequence of activities, we need the concept of the *time-slice* of an activity-instance. If s is an activity instance and $t^b_s \leq t_0 < t_1 \leq t^f_s$, $s \upharpoonright [t_0, t_1]$ denotes the time-slice of s that results from restricting s to the time interval $[t_0, t_1]$.

Many processes consist of activities performed on a particular object or objects. As with the distinction between activities and processes, the identification of a process instance as a sequence of activity instances acting on a particular object or objects is largely in the eyes of the beholder. Formally, an activity or process instance *acting on* an object a is simply an activity or process instance in which a is a constituent. For example, in the case of our example process PD, when a specific part a is fed into the paint–dry process, the process is said to *act on* a, and the process instance is said to be a process instance acting on a. The significance of this notion is that when we focus on the process as a process acting on a, each application of the paint, dry, queue, and examine activities involves the same object a as it cycles through the process loop. In an efficiently organized process, at any one time there will be several objects "in the pipeline," with one part being dried while the other is painted (or perhaps several parts going through the paint unit and several parts being dried at any one time). Our model has to distinguish between the various different process activities being applied to different objects and their sequential application to one particular object. For it is largely the latter case that we mean when we refer to a structured sequence of activities as a 'process'.

Since activities and processes are type-level objects, the formal definition of an activity or process acting on an object only makes sense in terms of parameters. Given a parameter $\dot{a}$ of the type of object on which an activity or process A can be applied, we write $A(\dot{a})$ to indicate that A is an activity or process acting on $\dot{a}$.

Our next step is to provide situation-theoretic machinery to model the way that two or more activities can be chained together to create a process.

Let $U(\dot{a})$ and $V(\dot{b})$ be processes acting on $\dot{a}$ and $\dot{b}$, respectively. Then the *fusion* of U, V is the process T such that, whenever $s : T$, then for some time t, $t^b_s < t < t^f_s$, we have $s \upharpoonright [t^b_s, t] : U$ and $s \upharpoonright [t, t^f_s] : V$. We denote this activity T by $U \otimes V$.

Note that if U is a process acting on $\dot{a}$ and V is a process acting on $\dot{b}$, then $U \otimes V$ is a process acting on $\dot{a}$ and $\dot{b}$. Often we want to combine U and V so that their fusion is acting on a single object. We do this with the following variant on the fusion operation.

Let $U(\dot{a})$ and $V(\dot{b})$ be processes acting on $\dot{a}$ and $\dot{b}$, respectively. Then the *fusion* of U, V *over* the relation $\dot{a}{=}\dot{b}$ is the process T such that, whenever $s : T$, then for some time t, $t^b_s < t < t^f_s$,

1. $s \upharpoonright [t^b_s, t] : U$ and $s \upharpoonright [t, t^f_s] : V$
2. $s \models \dot{a} = \dot{b}$

We denote this activity T by $U \otimes_{\dot{a}=\dot{b}} V$.

Note that $U \otimes_{\dot{a}=\dot{b}} V$ is a process acting on $\dot{a}$. All that this means is that $\dot{a}$ is a constituent of $U \otimes V$. Likewise, $U \otimes_{\dot{a}=\dot{b}} V$ is a process acting on $\dot{b}$. Suppose however that $s : U \otimes_{\dot{a}=\dot{b}} V$. Then for some anchor $f : \{\dot{a}, \dot{b}\} \to s$,

$$s : U \otimes_{\dot{a}=\dot{b}} V[f]$$

Since $s \models \dot{a} = \dot{b}$, we have $f(\dot{a}) = f(\dot{b}) = c$ for some object c in s. Thus, both $U[f]$ and $V[f]$ act on the same object c. Hence, regarded as a process on $\dot{a}$, $U \otimes_{\dot{a}=\dot{b}} V$ is a process in the desired, stronger sense that in any instantiation, both U and V act on the same object.

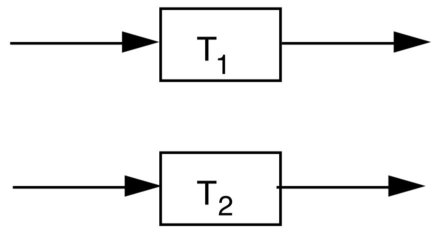


Figure 3: Parallel processes

More generally, a *process acting on* objects $a_1, \ldots, a_n$ is an activity in which $a_1, \ldots, a_n$ are constitutents. The objects $a_1, \ldots, a_n$ may be specific objects or parameters. We write $T(a_1, \ldots, a_n)$ to indicate that T is a process acting on $a_1, \ldots, a_n$.

An *instantiation* of a process T acting on $\dot{p}_1, \ldots, \dot{p}_r$ is an activity instance s such that for some anchor $f : \{\dot{p}_1, \ldots, \dot{p}_r\} \rightarrow s$, $s : T[f]$.

Suppose $U(\dot{p}_1, \ldots, \dot{p}_n)$ and $V(\dot{q}_1, \ldots, \dot{q}_m)$ are processes, and $R_1(\dot{p}_1, \ldots, \dot{p}_n, \dot{q}_1, \ldots, \dot{q}_m)$, ..., $R_k(\dot{p}_1, \ldots, \dot{p}_n, \dot{q}_1, \ldots, \dot{q}_m)$ are relations. Then the *fusion* of U, V *over* $R_1, \ldots, R_k$ is the activity, T, such that whenever $s : T$,

1. for some time t, $t_s^b < t < t_s^f$, $s \restriction [t_s^b, t] : U$ and $s \restriction [t, t_s^f] : V$
2. $s \models \ll R_1, \dot{p}_1, \ldots, \dot{p}_n, \dot{q}_1, \ldots, \dot{q}_m, 1 \gg \wedge \ldots \wedge \ll R_k, \dot{p}_1, \ldots, \dot{p}_n, \dot{q}_1, \ldots, \dot{q}_m, 1 \gg$

We denote the activity T by $U \otimes_{R_1,\ldots,R_k} V$.

Note that, if U is a process acting on $p_{i_1}, \ldots, p_{i_r}$ and the relations $\dot{p}_{i_1} = \dot{q}_{j_1}, \ldots, \dot{p}_{i_r} = \dot{q}_{j_r}$ are among $R_1, \ldots, R_k$, then in any instantiation in which $p_{i_1}, \ldots, p_{i_r}$ are anchored to specific objects $c_1, \ldots, c_r$, the instantiations of U and V both act on $c_1, \ldots, c_r$.

Let T be a process. The *iteration* of T is the process T^* such that, whenever $s : T^*$, then for some positive integer n and some $t_0 = t_s^b < t_1 < \ldots < t_n = t_s^f$, $s \restriction [t_i, t_{i+1}] : T$ for each $i = 0, \ldots, n-1$.

If T is a process acting on $\dot{a}$, then so too is T^*. (In the case of iteration, the same parameter $\dot{a}$ occurs in each instance of T in the loop, so the type T^* automatically captures the intuitive idea of the repeated application of the process T to the same object.)

If T_1 and T_2 are processes on $p_{i_1}, \ldots, p_{i_r}$, then so too is $T_1 \,\&\, T_2$, where $T_1 \,\&\, T_2$ is the type of all those activity instances s such that $s = s_1 \cup s_2$ where:

1. s_1 and s_2 have the same start and finish times;
2. $s_1 : T_1$ and $s_2 : T_2$.

$T_1 \& T_2$ is the process that consists of performing T_1 and T_2 simultaneously, in parallel, as illustrated in Figure 3. For example, Figure 4 illustrates a process where two differently shaped components are cleaned and then welded together. The entire process CW is of the form

$$[C_A \& C_B] \otimes W$$

Let's turn now to the activity "Examine part" in the process PD. In any given instantiation of PD, depending on the outcome of the "Examine part" activity, it will be followed by exactly one of the activities "Queue part" and "Release part." Thus, the process instantiation will involve exactly one of the two types

$$[\text{Examine part}] \otimes [\text{Queue part}]$$

$$[\text{Examine part}] \otimes [\text{Release part}]$$

So far, we do not have any situation-theoretic machinery to handle such a case.

Let $P(\dot{a})$ be any unary predicate. We say an activity $U(\dot{a})$ *resolves* P if, for any activity s and any constituent a of s such that $s : U[\dot{a}/a]$

$$\text{either} \quad s \models \ll P, a, t_s^f, 1 \gg \quad \text{or} \quad s \models \ll P, a, t_s^f, 0 \gg$$

where $U[\dot{a}/a]$ denotes the type $U[f]$ for the anchor $f = \{\langle a, \dot{a} \rangle\}$.

If $U(\dot{a})$ resolves $P(\dot{a})$ and if $V(\dot{b})$ and $W(\dot{c})$ are two activities, then

$$U \otimes_P \langle V, W \rangle$$

is the type of all activity instances s such that for some time t such that $t_s^b < t < t_s^f$:

1. $s \upharpoonright [t_s^b, t] : U[\dot{a}/a]$
2. if $s \models \ll P, a, t, 1 \gg$, then $s \upharpoonright [t, t_s^f] : V$ and $s \models \dot{a} = \dot{b}$
3. if $s \models \ll P, a, t, 0 \gg$, then $s \upharpoonright [t, t_s^f] : W$ and $s \models \dot{a} = \dot{c}$

We call $U \otimes_P \langle V, W \rangle$ the *branching fusion* of V, W with respect to P.

More generally, if $P_1(\dot{a}), \ldots, P_r(\dot{a})$ are mutually exclusive predicates such that

$$s \models \ll P_1, a, 1 \gg \vee \ldots \vee \ll P_r, a, 1 \gg$$

for all appropriate s, a, then we say $U(\dot{a})$ *resolves* $\langle P_1(\dot{a}), \ldots, P_r(\dot{a}) \rangle$ if, for any activity s and any constituent a of s such that $s : U[\dot{a}/a]$, we have

$$s \models \ll P_i, a, t_s^f, 1 \gg$$

for some (and hence exactly one) i.

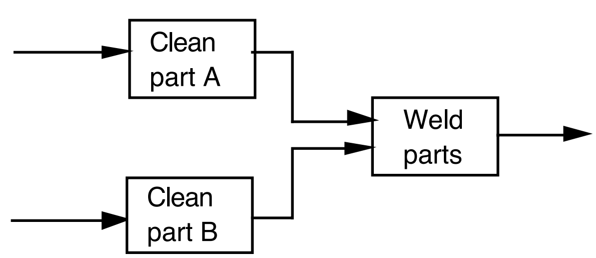


Figure 4: CW: The cleaning–welding process processes

If $U(\dot{a})$ resolves $\langle P_1(\dot{a}), \ldots, P_r(\dot{a})\rangle$ and if $V_1(\dot{b}_1), \ldots, V_r(\dot{b}_r)$ are activities, then

$$U \otimes_{P_1,\ldots,P_r} \langle V_1, \ldots, V_r\rangle$$

is the type of all activity instances s such that for some time t such that $t_s^b < t < t_s^f$:

1. $s \restriction [t_s^b, t] : U[\dot{a}/a]$
2. if $s \models \ll P_i, a, t, 1 \gg$, then $s \restriction [t, t_s^f] : V_i$ and $s \models \dot{a} = \dot{b}_i$

We call the process $U \otimes_{P_1,\ldots,P_r} \langle V_1, \ldots, V_r\rangle$ the *multi-branching fusion* of $U, V_1, \ldots, V_r$ with respect to $P_1, \ldots, P_r$.

Note that if U acts on $\dot{a}$, so too does $U \otimes_P \langle V_1, \ldots, V_r\rangle$.

Given an initial collection $\mathcal{A}$ of activities and parameters $\dot{p}_1, \ldots, \dot{p}_r$ for objects, the class $\mathcal{P}(\mathcal{A})$ of processes acting on $\dot{p}_1, \ldots, \dot{p}_r$ built on $\mathcal{A}$ is the smallest class $\mathcal{P}$ of processes acting on $\dot{p}_1, \ldots, \dot{p}_r$ such that:

1. $\mathcal{A} \subseteq \mathcal{P}$;
2. if $T_1, T_2 \in \mathcal{P}$, then $T_1 \otimes T_2 \in \mathcal{P}$;
3. if $T_1, T_2 \in \mathcal{P}$, then $T_1 \& T_2 \in \mathcal{P}$;
4. if T is a branching fusion of U, V_1, $\ldots$, $V_r \in \mathcal{P}$, then $T \in \mathcal{P}$;
5. if $T \in \mathcal{P}$, then $T^* \in \mathcal{P}$.

3 Sacks' baby: an application in sociolinguistics

[The work described in this section was done jointly with Duska Rosenberg and reported first in our paper [3].]

In the early 1990s, Rosenberg and I made use of insights from ethnomethodology to describe how speaker and listener cooperate to achieve shared understanding in the course of communicative, natural language interaction, using tools from situation theory to provide an ontological foundation for our analysis. We carried out this analysis as a first step towards the development of a situation-theoretic treatment of cooperative action, which we subsequently reported on in our joint monograph [4].

The focus of our analysis was a seminal paper in sociology by Harvey Sacks [7]. Sacks' paper is concerned with natural conversation, and in particular the way speaker and listener make use of social knowledge in the utterance and understanding of a simple natural language communication. He believes that the particular choice of words used by a speaker in, say, a description is critically influenced by the speaker's social knowledge, and that the listener utilizes his[1] social knowledge in order to interpret, in the manner the speaker intended, the juxtaposition of these words in conversation.

Note that this is quite different from the starting point taken in situation semantics. In situation semantics, context plays a significant role, but that context is very much a *linguistic* context, not a social one; situation semantics is, after all, a *semantic* theory, and its aim is to investigate the mechanisms by which the *meaning* of a natural language utterance enables that utterance to convey information, where 'meaning' itself becomes a formally defined notion within the theory.

[1] For convenience, we assume a female speaker and a male listener throughout the paper.

Rosenberg and I set out to see how situation theory can help formalize the arguments presented in [7] and to make use of that formalization to examine those arguments in a critical fashion.

The initial data Sacks works with consists of the following two opening sentences from a child's 'story':

The baby cried. The mommy picked it up.

As Sacks observes, when heard by a competent speaker of English, this is almost certainly heard as referring to a very small human (though the word 'baby' has other meanings in everyday speech) and to that baby's mommy (though there is no genitive in the second sentence, and it is certainly consistent for the mommy to be some other child's mother); moreover it is the baby that the mother picks up (though the 'it' in the second sentence could refer to some object other than the baby). Why do we almost certainly, and without seeming to give the matter any thought, choose this particular interpretation?

To continue, we are also likely to regard the second sentence as describing an action (the mommy picking up the baby) that follows, and is caused by, the action described by the first sentence (the baby crying), though there is no general rule to the effect the sentence order corresponds to temporal order or causality of events (though it often does so).

Moreover, we may form this interpretation without knowing what baby or what mommy is being talked of.

A further, more fundamental, observation Sacks raises at this point is that we recognize the utterance of these two sentences as a 'possible description'. As Sacks notes, it is of crucial importance for social scientists that there are notions such as 'possible descriptions', that are recognized by members of a society, without the need to examine the circumstances described, since those notions ('possible descriptions' and whatever) then constitute a collection of entities that the social scientist may legitimately study.

These then are the data and starting point for Sacks' discussion. As Sacks notes, the chosen example is extremely simple. But, he claims, far from rendering his study trivial, this very simplicity makes his observations all the more striking. He writes:

> My reason for having gone through the observations I have so far made was to give you some sense, right off, of the fine power of a culture. It does not, so to speak, merely fill brains in roughly the same way, it fills them *so that they are alike in fine detail*. The sentences we are considering are after all rather minor, and yet all of you, or many of you, hear just what I said you heard, and many of us are quite unacquainted with each other. I am, then, dealing with something *real* and something *finely powerful*. [Italics added.]

3.1 Ethnomethodology

At this point, I need to make clear the standpoint from which Sacks approaches his study. I shall do this from a situation-theoretic standpoint, structuring my account to be consistent with Chapter 4 of Suchman's book [8].

The traditional approach to social study is to commence with a foundational framework consisting of a collection of empirically identified social norms. This approach is sometimes referred to as 'normative sociology'. Thus normative sociology posits, and then attempts to describe, an objective world of social facts, or received norms, to which our attitudes and actions are a response.[2] What the everyday practice of normative sociology then amounts to is constructing a social science by a process of refinement, or quantification, of an initial collection of foundational norms that constitute our common-sense-view of the world. This enterprise is successful in so far as it improves upon (in some way) the normative structure it starts with. Human action is explained by reference to the body of social norms taken to be the grounding structure for the theory.

Sacks rejected this approach to social science. Instead, he took a radically different stance known as *ethnomethodology.* Rather than take a normative social structure as foundational and explain human action in terms of those norms, the ethnomethodologist regards human action as fundamental and seeks to explain *how* human action can give rise to what we perceive as a collection of norms. Thus for the ethnomethodologist, the common-sense-view of the world is taken not as a foundational structure to be improved upon but as the *topic* to be studied. Instead of assuming that we act in response to an objectively given social world, we assume that our everyday social practices render the world publicly available and mutually intelligible [8, p. 57]. The mutual intelligibility and objectivity of the social world that traditional social science treats as a *received*, conceptual scheme or collection of norms to govern action, is regarded instead as *derivative* on social activity, and the issue to study is how social action gives rise to this state of affairs. The common sense of the social world that traditional studies take as a precondition for interaction becomes instead its product, and the objective reality of social facts are no longer the fundamental *principle* of social science but that science's fundamental *phenomenon.* Social structure is thus seen as an emergent product of action, not its foundation.

So, rather than look for structures that are invariant across situations, what Sacks does in his paper is investigate the processes by which particular interactions result in shared meaning and rational accountability. As Suchman says [8, pp.50–51]

> ... as a consequence of the indexicality of language, mutual intelligibility is achieved on each occasion of interaction with reference to situation particulars, rather than being discharged once and for all by a stable body of shared meanings.

3.2 Sacks' analysis

I summarize Sacks' analysis of the two sentences. In the subsequent section I shall re-cast his argument in situation-theoretic terms and investigate the demands this places on various members of the ontology of situation theory.

Sacks first of all introduces what he calls *categories.* In situation-theoretic terminology, these are certain kinds of object-types.

For example, the following are categories:

$$\begin{aligned}
\text{'male'} &= T_{male} &&= [\dot{p} \mid w \models \ll \text{male}, \dot{p}, t_{now}, 1 \gg];\\
\text{'female'} &= T_{female} &&= [\dot{p} \mid w \models \ll \text{female}, \dot{p}, t_{now}, 1 \gg];\\
\text{'baby'} &= T_{baby} &&= [\dot{p} \mid w \models \ll \text{baby}, \dot{p}, t_{now}, 1 \gg];\\
\text{'mommy'} &= T_{mother} &&= [\dot{p} \mid w \models \ll \text{mother}, \dot{p}, t_{now}, 1 \gg];
\end{aligned}$$

where $\dot{p}$ is a parameter for a person.

In the above type definitions, the situation w is 'the world', by which we mean any situation big enough to include everything under discussion.

[2] Paraphrased from [8, p. 54], where Suchman argues convincingly against such an approach.

Of course, the above correspondences between Sacks' categories and situation-types are not in any sense reductionist definitions. Rather we are just translating Sacks' notions (category 'male', category 'female', etc.) into ours (type T_{male}, type T_{female}, etc.). In situation theory we start with a collection of relations provided by the individuation scheme of the agent–species, and use them to determine types. Thus, although the above *types* are defined within the theory, the basic *properties* of being a person, being male, being female, are initial *givens* — just as the corresponding categories are for Sacks' analysis.

Sacks then defines a *(membership) categorization device* to be a non-empty collection of categories, together with rules of application. For example, the categorization device 'gender'. This device consists of the two categories 'male' and 'female', together with the rule for applying these categories to (say) human populations. Other examples are the 'family' categorization device, which consists of categories such as 'baby', 'mommy', 'daddy', etc. and the 'stage-of-life' device, which consists of categories such as 'baby', 'child', 'adult', etc. (We shall presently examine the appearance of the same category in two or more devices, as occurs here with the 'baby' category.)

According to Sacks, it is important that the categories in a device 'go together' (his phrase). In the situation-theoretic treatment, this requirement is met automatically, and elegantly, since both categories and categorization devices are captured using the same concept of types. That is, the collections of categories in categorization devices correspond to further types. In the case of a categorization device, we need to endow the corresponding type with some structure in order to obtain an acceptable analogue of what Sacks means by a device.

For instance, the devices 'gender, 'family', and 'stage-of-life' correspond to the respective types:

$$\begin{aligned}
T_{gender} &= [\dot{e} \mid w \models \ll =, \dot{e}, T_{male}, t_{now}, 1 \gg \vee \ll =, \dot{e}, T_{female}, t_{now}, 1 \gg];\\
T_{family} &= [\dot{e} \mid w \models \ll \text{family}, \dot{e}, t_{now}, 1 \gg];\\
T_{stage\text{-}of\text{-}life} &= [\dot{e} \mid w \models \ll \text{stage-of-life}, \dot{e}, t_{now}, 1 \gg];
\end{aligned}$$

where $\dot{e}$ is a parameter for a type.

Thus, T_{gender} is the type of all those types $\dot{e}$ such that (in the world) $\dot{e}$ is either the type T_{male} or the type T_{female}. In other words, T_{gender} is the type that itself consists of just two types, the type T_{male} and the type T_{female}.

Note that, whereas the type T_{gender} can be described extensionally (as we have just described it), the same is not the case for the types T_{family} and $T_{stage\text{-}of\text{-}life}$. Though in the case of these particular two examples it is possible to give extensional definitions, there are significant problems associated with such a move, since the notions involved are intrinsically intensional.[3]

As examples of the rules of application that are part of a categorization device, Sacks gives the following. First, *the economy rule*, a reference adequacy rule that says:

(ER) A single category from any device can be referentially adequate.

[3] Exactly what is meant by the words 'intensional' and 'extensional' is not easily pinned down, apart from the fact that each is the complement of the other. A property of extensionality that can serve as a characterization for our present purposes is that if a type T is extensional, then the appropriate domain in the universe can be seen by an agent to divide cleanly into two collections, those that are of type T and those that are not of type T. Thus the collection of all persons divides pretty clearly into the males and the females (type T_{gender} is extensional) but it is not at all clear how to obtain a clean classification of a given society into families—exactly what constitutes a 'family' is both context-dependent and vague (type T_{family} is intensional). We shall have more to say on the subject of types in due course.

For instance, the economy rule allows use of the phrase ‘the baby’ to be referentially adequate — see presently for a more complete picture.

Sacks’ second rule, *the consistency rule*, is a relevance rule that says:

> (CR) If some population of persons is being categorized, and if a category from some device has been used to categorize one member of the population, then that category, or other categories of the same device, may be used to categorize further members of the population.

For instance, the device ‘family’ having been used to refer to some baby by means of the category ‘baby’, further persons may be referred to by other categories in the same device, such as ‘mommy’ and ‘daddy’. Again, we shall present a more complete picture in due course.

Associated with the consistency rule, Sacks formulates the following ‘hearer’s maxim’:

> (HM1) If two or more categories are used to categorize two or more members of some population, and those categories can be heard as categories from the same device, then hear them that way.

For instance it is in this way that in the two sentences under consideration, ‘baby’ and ‘mommy’ are heard as from the ‘family’ device. But notice that this does not preclude our simultaneously hearing ‘baby’ as from the stage-of-life device—indeed, as Sacks himself argues, this is probably what does occur.

Now, the above hearer’s maxim does not fully capture what goes on in the example under consideration. For it is not just that ‘baby’ and ‘mommy’ are heard as belonging to the same category ‘family’; rather they are heard as referring to individuals in *the very same* family. Sacks explains this by observing that the device ‘family’ is one that is, what he calls *duplicatively organized*.[4] He explains this notion as follows:

> When such a device is used on a population, what is done is to take its categories, treat the set of categories as defining a unit, and place members of the population into cases of the unit. If a population is so treated and is then counted, one counts not numbers of daddies, numbers of mommies, and numbers of babies but numbers of families—numbers of ‘whole families’, numbers of ‘families without fathers’, etc. A population so treated is partitioned into cases of the unit, cases for which what properly holds is that the various persons partitioned into any case are ‘coincumbents’ of that case.

The following hearer’s maxim is associated with duplicatively organized devices:

> (HM2) If some population has been categorized by means of a duplicatively organized device, and a member is presented with a categorized population which can be heard as ‘coincumbents’ of a case of that device’s unit, then hear it that way.

According to Sacks, it is this maxim that results in our hearing ‘the baby’ and ‘the mommy’ in our example as referring to individuals in the very same family.

Sacks next point is that the phrase ‘the baby’ is in fact heard not just in terms of the ‘family’ device but simultaneously as from the ‘stage-of-life’ device. The reason is, he claims,

[4] The notion of duplicative organization plays a significant role in our situation-theoretic treatment, where it corresponds to a certain kind of structure on constraints.

that ‘cry’ is, in his terminology, a *category-bound* activity, being bound to the category ‘baby’ in the stage-of-life device. We shall postpone until later a discussion of the argument Sacks puts forward in favor of this notion, and the further elaboration he provides, noting only that ‘cry’ is just one of many instances of activities that are bound in this way to a particular category or categories in a particular device.

Sacks codifies what it is that leads us to hear ‘baby’ as from the ‘stage-of-life’ device in addition to the ‘family’ device, by means of a further hearer’s maxim:

> (HM3) If a category-bound activity is asserted to have been done by a member of some category where, if that category is ambiguous (i.e. is a member of at least two devices), but where, at least for one of those devices, the asserted activity is category-bound to the given category, then hear at least the category from the device to which it is bound.

The final part of Sacks’ analysis that we shall consider here concerns the way that an observer describes a particular scene. For instance, if you were to observe a very small human crying, you would most likely describe what you saw as “A baby is crying”, or some minor variant thereof. You are far less likely to say “A person is crying” or, even if you could identify the gender of the baby as female, “A girl is crying”. Again, if you subsequently saw a woman pick up that baby, and if that woman looked about the age to be the baby’s mother, you would probably describe what you saw as “It’s mother picked it up”. You are less likely to say “A woman picked it up”, and even less likely still to say “A person picked it up”. What you would see, and what you would so describe, is that a *baby* cried and *it’s mother* picked it up. Moreover, you would most likely take it that the mother picked up the baby *because* it cried, and that her intention was to *comfort* the child. (Actually, this example is a bit more subtle than we have indicated. Most likely your description would be influenced by *the way* the woman behaved towards the baby. But if she *acts* like a mother, that is probably how you would describe her, even though several other descriptions would be ‘accurate’.)

Why is this? Well, Sacks explains the first of these observations, the use of the phrase ‘the baby’, by means of the following ‘viewer’s maxim’:

> (VM1) If a member sees a category-bound activity being done by a member of a category to which the activity is bound, then see it that way.

Since the activity of crying is category-bound to the category ‘baby’ in the stage-of-life device, this is the natural way to see and to describe the activity, whenever such a way of seeing and describing is possible.

Turning to the remaining set of observations, there are social norms that govern, or can be seen to govern, the actions of members of the society, and one such norm is that a mother will comfort her crying baby. This is a very powerful social norm, and society demonstrates strong disapproval for a mother who fails to conform to it. Sacks’ point now is that in addition to governing behavior, where by ‘governing’ we may mean nothing more than that the norm serves to describe a normal way of behaving, norms fulfill a further role; namely, viewers use norms to provide some of the orderliness of the activities they observe. In this case, we may capture such a use of a norm by means of a second viewer’s maxim:

> (VM2) If one sees a pair of actions which can be related via a norm that provides for the second given the first, where the doers can be seen as members of the categories the norm provides as proper for that pair of actions, then (a) see that the doers are such members, and (b) see the second as done in conformity with the norm.

By means of VM2(a), the viewer sees the person who picks up the baby as the baby's mother, provided it is possible to see it thus, and moreover, by VM2(b), takes it that this action is performed in accordance with the norm that says that mothers comfort their crying babies.

So much then for the initial part of Sacks' analysis.[5] I turn now to the aim of recasting his arguments in situation-theoretic terminology, using the ontology and tools of situation theory.

3.3 A situation-theoretic analysis

I start by examining Sacks' explanation as to why the hearer hears 'the baby' as referring to a member of a category in the 'family' device and 'the mommy' as a member of the same category, and indeed the same unit within that category. Now at this point, Sacks' account is not completely clear as to when he is talking about what enables the speaker to utilize the various words she does, and which points concern the hearer. This, however, makes a difference to the analysis, as we indicate presently.

In the case of the adequate use the phrase 'the baby' to refer to a member of the category 'baby' in the 'family' device, it seems that the performance of both speaker and hearer may be explained by the referential adequacy rule ER: *a single category from any device can be referentially adequate.*

Sacks appears to have formulated this rule with the hearer in mind. In addition to his having declared at some length that his main concern is how hearer's hear things a certain way, when he comes to formulate the ER rule, he says [emphasis added]:

> ...It is not necessary that some multiple of categories from categorization devices be employed *for recognition that* a person is being referred to, to be made; a single category will do. (I do not mean by this that more cannot be used, only that for reference to persons *to be recognized*, more need not be used.)

In our analysis, however, we shall concentrate on both speaker and listener, as we seek to describe the mechanisms they invoke to achieve successful communication. Of course, since successful communication depends on both speaker and listener cooperating, none of the rules Sacks introduces will concern only one party. However, one of the advantages that is gained by including the information flow as part of our study, which is what is done in situation theory, is that we are able to pull apart the speaker and listener actions, and track the manner in which the speaker invokes mechanisms that enable the listener to correctly interpret the utterance.

In point of fact, we believe that the rule ER, as stated, serves both parties. But what about the referential adequacy of the phrase 'the mommy' in the second sentence? Here Sacks invokes a combination of ER together with his relevance rule CR, which says, in part [emphasis added]:

> ...if a category from some device *has been used* to categorize one member of the population, *then* that category, or other categories of the same device, may be used to categorize further members of the population.

Now for the speaker, this seems fine. But it does not explain how it is that the hearer hears it in the appropriate way. Certainly, the utterance of the phrase 'the baby', or even the entire first sentence, is not sufficient to bring the 'family' device into sufficient focus to provide adequate reference, via ER, to 'the mommy'. Rather the *combination* of the two phrases 'the baby' and 'the mommy' seems to be what is required to trigger the appropriate structures in the hearer.

[5] In [7], the remaining discussion concerns the reasons why the two sentences considered may be regarded as constituting a 'story'.

For example, what if the two sentences spoken were one of the following two alternatives?

The baby cried. The aunt picked it up.

The baby cried. The doctor picked it up.

In the first of these examples, there is certainly a use of the 'family' device, but in this case the family unit invoked is not the 'father–mother–children' unit involved, surely, in the Sacks example. A quite different notion of a 'family' is used here, the 'extended-family', and it is the second sentence that serves to determine this unit.

Turning to the doctor example, this does not even invoke the 'family' device at all; the second sentence here serves to invoke a 'medical care' device, whose categories are 'doctors', 'nurses', 'surgeons', 'patients', and the like.

In general, it is then the *combined* effect of all parts of an utterance that serves to evoke the various devices used to provide the appropriate interpretations of the words uttered. Though he does not make this point himself, Sacks' hearer's maxim HM1 is certainly consistent with this observation: *If two or more categories are used to categorize two or more members of some population, and those categories can be heard as categories from the same device, then hear them that way.*

With this clarification, Rosenberg and I found as reasonable Sacks' explanation as to why 'the baby' and 'the mommy' are heard as being in the 'family' device. We were less happy with his use of HM2 to explain why these two words are heard as referring to individuals in the same family unit. One problem was the one just considered, namely that the entire utterance may be required in order to determine the appropriate unit, in this case the 'close-family' unit consisting, in typical cases, of a mother, a father, and their children. But this is a relatively minor issue involving the circumstances under which HM2 may be applied, and is easily corrected. Far more problematical is the overall validity of HM2.

Suppose that instead of the two sentences Sacks considers, the child had uttered the following:

The baby cried. My mommy picked it up.

Now it is certainly possible that both 'the baby' and 'my mommy' *can be* heard as referring to coincumbents of the same family unit (when the utterance is made by the baby's sibling), but this is not the way this utterance would normally be heard—quite the contrary in fact. In this case, therefore, HM2 would seem to lead to the entirely wrong conclusion. Simply because something can be heard a certain way does not necessarily imply that it will or should be heard that way. Probably the strongest valid observation that can be made here is that things *should* be heard the way they are *normally* heard, which is verging on the trivially tautological, but seems to say something for all that. Thus, we propose to replace Sacks' HM2 by the following maxim:

> (HM2$'$) If some population has been categorized by means of a duplicatively organized device, and a member is presented with a categorized population which would normally be heard as coincumbents of a case of that device's unit, then hear it that way.

In fact, when we come to reformulate Sacks' argument in situation-theoretic terms, we shall avoid this difficulty. We provide a constraint structure on types that corresponds to duplicative organization, and when we work directly with this structure, we seem to obtain just the right conclusions, without the need for making an assumption such as HM2$'$.

A similar problem arises with Sacks' HM1 and with the viewer's maxim VM2, but these do not play a crucial role in Sacks' own analysis of the particular data.

Note that the situation-theoretic analysis assumes that the speaker's perspective has been determined. That is, we do not ask ourselves *why* the speaker chooses the particular form of words she does, an issue closely related to the question why we see things in a certain way, but rather we use the framework of situation theory to track the way the speaker and listener cooperate in order for the communicative act to be successful.

One immediate question concerns the use of the definite noun phrases 'the baby' and 'the mommy'. Use of the definite article generally entails uniqueness of the referent. In the case of the phrase 'the baby', where, as in the Sacks example, no baby has previously been introduced, one would normally expect this to be part of a more complex descriptive phrase, such as 'the baby of the duchess's maid', or 'the baby on last night's midnight movie'. So just what is it that enables the speaker to open an explanation with the sentence 'The baby cried'? It could be argued that an implicit suggestion for an answer lies in his later discussion of proper openings for 'stories', but this is a part of his article we do not consider here.

For a situation-theoretic analysis, there is no problem here. The situation theorist assumes that all communicative acts involve a *described situation*, that part of the world the act is *about*. Exactly how this described situation is determined varies very much from case to case. For example, the speaker may have witnessed, read about, or even imagined the event she describes. In the Sacks case, the speaker *imagines* a situation in which a baby cried and its mother picked it up. Let s denote that situation.

Notice that although s has a definitive ontological status within the situation-theoretic world — it is a definite situation for which we introduce a specific name, s — it does not follow that s has an *extensional definiteness* for either speaker or listener. Though in typical cases the speaker may have extensive knowledge of s, she is unlikely to know *everything* there is to know about s, and in all likelihood can only specify it in vague, general terms, such as 'the scene I witnessed' or 'the situation I read about'. As for the listener, unless he witnessed the same scene or read the same book (say), he is likely to have even less information about s. For the listener, s may simply be 'the situation the speaker is currently telling me about'. Nevertheless, for all the vagueness concerning the identity of described situations, speakers and listeners do successfully communicate, on a widespread and regular basis, and moreover that communication involves the conveyance of information about some part or aspect of the world, i.e. a situation. Put loosely, successful communication involves all parties being aware that they are talking and listening *about something*, and moreover that they are talking and listening about *the same thing*. Though each participant may have vastly different kinds and amounts of information about that 'same thing', and have different ways of describing what 'it' is, ways that will evolve during the course of the communication, this does not prevent the situation theorist from taking the described situation as a definite element in the ontology. It simply means that the ensuing theory will have to allow for, and hopefully account for, the vague and partial nature of the information about that situation possessed by each participant in the communicative act.

In the Sacks example, the described situation s will be such that it involves one and only one baby, otherwise the use of the phrase 'the baby' would not be appropriate. In starting a communicative act with the sentence 'The baby cried', the speaker is informing the listener that she is commencing a description of a situation, s, in which there is exactly one baby, call it b. (Whether or not b is a real individual in the world, or some fictional entity, depends on s.

This does not affect the way our analysis proceeds.)

The *propositional content* of the utterance of the first sentence 'The baby cried', which in the case of a simple description like the present example is, for speaker and listener, the principal item of information about the described situation that is conveyed, is

$$s \models \ll \text{cries}, b, t_0, 1 \gg$$

where t_0 is the time, prior to the time of utterance, at which the crying took place. In words, in the situation s, the baby b was crying at the time t_0.

Notice that, in the absence of any additional information, the only means available to the listener to identify b is as the referent for the utterance of the phrase 'the baby'. The utterance of this phrase tells the listener two pertinent things about s and b:

$$b : T_{baby} \quad \text{(i.e. } b \text{ is of type } T_{baby}\text{)} \tag{1}$$

where T_{baby} is the type of all babies, and

$$b \text{ is the unique individual of this type in } s. \tag{2}$$

Now let's consider what information is conveyed, in addition to the propositional content, by the utterance of the entire first sentence 'The baby cried.'

As Sacks observes, the activity of crying is category bound to babies in the stage-of-life device, so the listener will (HM3) hear the utterance in such a way that

$$T_{baby} : T_{stage\text{-}of\text{-}life}. \tag{3}$$

That is to say, this item of information will be available to the listener as he processes the incoming utterance, and will influence the way the input is interpreted. (More precisely, the listener's cognitive state will be consistent with this item of information being salient, and we may describe the way he interprets what he hears in terms of such influence.)

We carry over Sacks' economy rule ER to our treatment, reformulating it in situation-theoretic terms as:

> ($\mathrm{ER}_{\mathrm{ST}}$) A single category (i.e. object-type) from any device (i.e. type of object-type) is adequate to determine an object that is uniquely of that category in the described situation.[6]

So far we have considered only the first sentence of the utterance. Before we look at the entire two-sentence utterance, where the speaker goes on to say 'The mommy picked it up', we note that the properties and types in the situation-theoretic ontology form a structured collection; there are relationships between various properties, between various types, and between types and properties. For instance, the following pairwise-related properties and types are all in, or

[6] Or in some other, resource situation. This alternative does not arise in the cases considered in this paper.

relevant to, the family device, T_{family} :

$M(x)$	the property of x being a mother
$B(x)$	the property of x being a baby
$M(x,y)$	the relation of x being the mother of y
T_{mother}	the type of being a mother
T_{baby}	the type of being a baby
$T_{mother\text{-}of}$	the 2-type that relates mothers to their offspring

What Sacks refers to as the 'duplicatively organized' nature of the family device, T_{family}, reflects itself in this type structure, in that the type $T_{mother\text{-}of}$ acts as a 'fundamental' one, with the types T_{mother} and T_{baby} being linked to, and potentially derivative on, that type. More precisely, the following structural constraints[7] are salient in the device T_{family}:

$$\begin{aligned} T_{mother} &\Rightarrow \exists \dot{y} T_{mother\text{-}of} \\ T_{baby} &\Rightarrow \exists \dot{x} T_{mother\text{-}of} \end{aligned}$$

where

$$T_{mother} = [\dot{x}, \dot{y} \mid w \models \ll \text{mother-of}, \dot{x}, \dot{y}, t_{now}, 1 \gg].$$

What does these mean? Well $T_{mother\text{-}of}$ is a 2-type, the type of all pairs of individuals x, y such that x is the mother of y (at the present time, in the world). The first of the above two constraints says that the type T_{mother} *involves* (or is *linked to*) the type $\exists \dot{y} T_{mother\text{-}of}$. This has the following consequence: in the case where $T_{mother} : T_{family}$ (i.e. T_{mother} is of type T_{family}) and $T_{baby} : T_{family}$, the following implications are salient:

$$p : T_{mother} \;\rightarrow\; \exists q\, (p, q : T_{mother\text{-}of}) \tag{4}$$

$$q : T_{baby} \;\rightarrow\; \exists p\, (p, q : T_{mother\text{-}of}). \tag{5}$$

These two implications are not constraints. In fact they do not have any formal significance in situation theory. They are purely guides to the reader as to where this is all leading. (4) says that if p is of type T_{mother} (i.e. if p is a mother), then there is an individual q such that the pair p, q is of type $T_{mother\text{-}of}$ (i.e. such that p is the mother of q). The salience of this implication for an agent $\mathcal{A}$ has the consequence that, if $\mathcal{A}$ recognizes that p is a mother then $\mathcal{A}$ will, if possible, look for an individual q of which p is the mother.

To continue with our analysis, as in the case of 'the baby', in order for the speaker to make appropriate and informative use of the phrase 'the mommy', the described situation s must contain exactly one individual m who is a mother. In fact we can make a stronger claim: the individual m is the mother of the baby b referred to in the first sentence. For if m were the mother not of b but of some other baby, then the appropriate form of reference would be 'a mother', even in the case were m was the unique mother in s. We can describe the mechanism that produces this interpretation as follows.

Having heard the phrase 'the baby' in the first sentence and 'the mommy' in the second, the following two items of information are salient to the listener:

$$m : T_{mother} \tag{6}$$

[7]The notion of constraint used here is a new one, which extends that described in [1].

$$m \text{ is the unique individual of this type in } s. \tag{7}$$

In addition, we shall show that the following, third item of information is also salient:

$$m \text{ is the mother of } b. \tag{8}$$

Following the utterance of the first sentence, the listener's cognitive state is such that the type T_{baby} is of type $T_{stage\text{-}of\text{-}life}$. This type has categories that include T_{baby}, T_{child}, $T_{adolescent}$, T_{adult}, all of which have equal ontological status within this device, with none being derivative on any other. But as soon as the phrase 'the mommy' is heard, the combination of 'baby' and 'mommy' switches the emphasis from the type $T_{stage\text{-}of\text{-}life}$ to the type T_{family}, making salient the following propositions:

$$T_{baby} : T_{family}. \tag{9}$$

$$T_{mommy} : T_{family}. \tag{10}$$

In the T_{family} device, the various family relationships that bind a family together (and which therefore serve to give this type its status as a type) are more fundamental than the categories they give rise to. In particular, the types T_{baby} and T_{mother} are derivative on the type $T_{mother\text{-}of}$ that relates mothers to their babies.

Now, proposition (9) is the precondition for the salience of implication (5), namely

$$q : T_{baby} \;\rightarrow\; \exists p\,(p, q : T_{mother\text{-}of}).$$

Substituting the particular individual b for the variable q, we get

$$b : T_{baby} \;\rightarrow\; \exists p\,(p, b : T_{mother\text{-}of}).$$

But by (1), we know that

$$b : T_{baby}.$$

Thus we have the salient information

$$\text{there is an } m \text{ such that } m, b : T_{mother\text{-}of}. \tag{11}$$

The use of the definite article in the phrase 'the mommy' then makes it natural to take this phrase to refer to the unique m that satisfies (11). Thus the listener naturally takes the phrase 'the mommy' to refer to the baby's mother. This interpretation is reinforced by the completion of the second sentence '...picked it up', since there is a social norm to the effect that a mother picks up and comforts her crying baby. This explains how the fact (8) becomes salient to the listener.

It should be noticed that the switch from the salience of one set of constraints to another was caused by the second level of types in the hierarchy. The constraints we were primarily interested in concerned the types T_{mother} and T_{baby}. These types are part of a complex network of inter-relationships (constraints). Just which constraints in this network are salient to the agent is governed by the way the agent encounters the types, that is to say, by the type(s) of those types—for instance, whether T_{baby} is regarded (or encountered) as of type $T_{stage\text{-}of\text{-}life}$ or of type T_{family}. This consideration of the type of the types linked by a constraint represents a much finer treatment of constraints and inference than has hitherto arisen in situation theory. By moving to a second level of typing (i.e. to types of types), we are able to track the way agents use one set of constraints rather than another, and switch from one set to another. The first level of types allows us to capture the informational connections between two objects; the second level allows us to capture the agent's preference of a particular informational connection, and thereby provides a formal mechanism for describing normality.

Of course, salience is not an all-or-nothing state of affairs. Preferential readings, interpretations, or actions can be over-ruled or ignored. An analogy that we find helpful is to think of salience as analogous to a gravitational field. At any point, the field pulls you in one particular direction, and depending on your circumstances you will follow that pull to a greater or lesser degree. If there are no other forces (constraints, influences) acting on you, you will follow the field in the direction of strongest pull. When there are other forces acting on you, you will still be under the influence of the field, but may move in a different direction, a direction that results from the combined effect of the gravitational field and those other forces. Normality can, then, be thought of as following the attraction of a gravitational field.

With our situation-theoretic analysis now complete, we look back and see how it relates to Sacks' treatment.

First of all, his rule ER corresponds to our $\mathrm{ER}_{\mathrm{ST}}$.

We have already taken issue with his rule CR. Appropriately modified as suggested earlier, this rule would say that once a particular type has been made salient, it may be used to categorize more than one individual. This seems reasonable enough, and is certainly what happens with our analysis.

Sacks' HM1 plays no crucial role in Sacks' analysis of the data, and hence does not really bear on our account, but we have already noted that we see a problem with this maxim.

Turning to the issue of duplicative organization and the hearer's maxim HM2, our treatment provides structural significance to Sacks' notion of a duplicatively organized device, in that such a device has one or more prominent types on which the others are potentially derivative. We found difficulty with Sacks' formulation and use of the hearer's maxim HM2. Concerning the use of this maxim, we observed that the conjunctive hearing of the two phrases 'the baby' and 'the mommy' is required to bring the duplicatively organized type T_{family} into salience, and thereby place emphasis on the type $T_{mother\text{-}of}$ among the various, relevant types derivative on it. In addition, in order to avoid a significant counterexample, we had to replace the maxim HM2 by a much weaker principle HM2$'$.

HM3 is consistent with our account, as are the viewer's maxims VM1 and VM2.

References

[1] Devlin, K. *Logic and Information*, Cambridge University Press (1991).

[2] Devlin, K. Situation Theory and the Design of Interactive Information Systems, in *Design Issues for* CSCW, Springer-Verlag (1994), pp.61–87.

[3] *Situation Theory and Cooperative Action* (with D. Rosenberg), in Aczel, P. Israel, D, Katagiri, Y. and Peters, S. (editors) *Situation Theory and its Applications, Volume 3*, CSLI Lecture Notes 37 (1993), pp.213–264.

[4] Devlin, K and Rosenberg, D. *Language at Work: Analyzing Communication Breakdown in the Workplace to Inform System Design*, Stanford, CA: CSLI Publications (1996).

[5] Menzel, C. and Mayer, R. Situations and Processes, *Concurrent Engineering: Research and Applications* 4(3) (1996).

[6] Menzel, C. and Mayer, R. A Situation Theoretic Approach to the Representation of Processes, in P. Bernus and L. Nemes (eds.), *Modelling and Methodologies for Enterprise Integration*, London, Chapman and Hall, 1996.

[7] Sacks, H. On the Analyzability of Stories by Children, in Gumpertz, J. & Hymes, D. (eds): *Directions in Sociolinguistics, The Ethnography of Communication*, Holt, Rinehart and Winston Inc. (1972), pp.325–345.

[8] Suchman, L. *Plans and Situated Actions: The Problem of Human Machine Communication*, Cambridge University Press (1987).