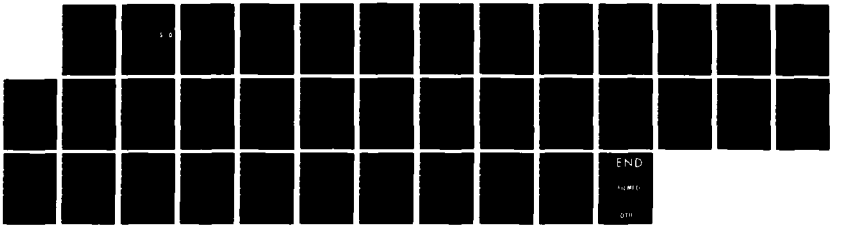
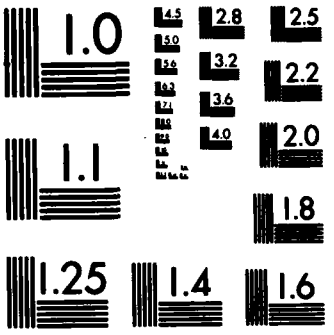
AD-A154 162 EXPLANATION: A FIRST PASS(U) YALE UNIV NEW HAVEN CT 1/1
DEPT OF COMPUTER SCIENCE R C SCHANK AUG 84
YALEU/CSD/RR-330 N00014-75-C-1111

UNCLASSIFIED F/G 5/10 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Explanation: A First Pass

Roger Schank

YALEU/CSD/RR #330

September 1984

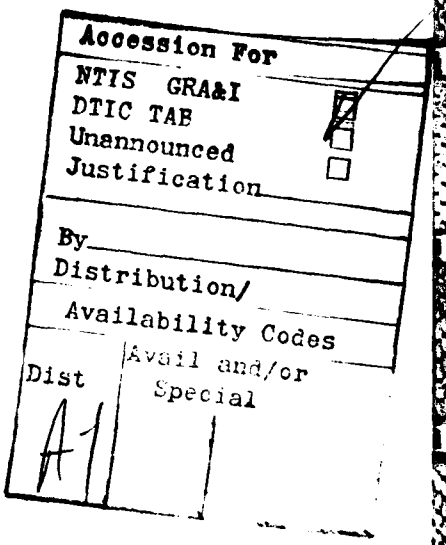
YALE UNIVERSITY
DEPARTMENT OF COMPUTER SCIENCE

EXPLANATION: A FIRST PASS

Roger C. Schank

YALEU/CSD/RR# 330

September 1984



This work was supported in part by the Advanced Research Projects Agency of the Department of Defense and monitored under the Office of Naval Research under contract N00014-75-C-1111 and contract N00014-82-K-0149, National Science Foundation IST-8120451, and the Air Force contract F49620-82-K0010.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
390	A154162	

4. TITLE (and Subtitle)

EXPLANATION: The First Pass

5. TYPE OF REPORT & PERIOD COVERED

Technical

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)

Roger C. Schank

8. CONTRACT OR GRANT NUMBER(s)

N00014-75-C-1111
N00014-82-K-0149

9. PERFORMING ORGANIZATION NAME AND ADDRESS

Yale University
Computer Science Department
10 Hillhouse Avenue, New Haven, CT 06520

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS

Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

12. REPORT DATE

August 1984

13. NUMBER OF PAGES

28

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

Office of Naval Research
Information Systems Program
Arlington, VA 22217

15. SECURITY CLASS. (of this report)

Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this report is unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Explanation
Understanding
Reminding
Memory

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

People explain their actions to others, and to themselves, every day. Why do we need to send and receive these explanations? What is the value of such explanations? After we explain a thing to ourselves or accept the explanation of another, what becomes of it? What is the point of these explanations? What is their role in the learning process? What do they tell us about what it means to be intelligent?

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

For some people, explanations do not play a big role. Many people are willing to observe events that would disturb others, and attribute these events to inexplicable circumstances. They may not even wonder about them at all. What is the difference between people who search for explanations for everything and those who do not require them? What are they doing differently? Is curiosity the major factor with little more significance than the entertainment value of the explanation? Is there some emotional satisfaction derived from knowing why things have happened the way they did, or, is something more significant afoot? Keywords: Understanding, Reminding, and Memory.

In this paper, I will discuss some of the issues surrounding the problem of explanation. Mostly, I will concentrate on making clear exactly what I believe the problem to be, and what issues need to be addressed in order to solve the problem. I will do this by concentrating on a number of examples that serve to illustrate the issues. I will only suggest an outline of some solutions however. Full solutions await detailed research.

OFFICIAL DISTIRUBTION LIST

Defense Documentation Center — 12 copies
Cameron Station
Alexandria, Virginia 22314

Office of Naval Research — 2 copies
Information Systems Program
Code 437
Arlington, Virginia 22217

Dr. Judith Daly — 3 copies
Advanced Research Projects Agency
Cybernetics Technology Office
1400 Wilson Boulevard
Arlington, Virginia 22209

Office of Naval Research — 1 copy
Branch Office - Boston
495 Summer Street
Boston, Massachusetts 02210

Office of Naval Research — 1 copy
Branch Office - Chicago
536 South Clark Street
Chicago, Illinois 60615

Office of Naval Research — 1 copy
Branch Office - Pasadena
1030 East Green Street
Pasadena, California 91106

Mr. Steven Wong — 1 copy
New York Area Office
715 Broadway - 5th Floor
New York, New York 10003

Naval Research Laboratory — 6 copies
Technical Information Division
Code 2627
Washington, D.C. 20375

Dr. A.L. Slafkosky — 1 copy
Commandant of the Marine Corps
Code RD-1
Washington, D.C. 20380

Office of Naval Research — 1 copy
Code 455
Arlington, Virginia 22217

Office of Naval Research — 1 copy
Code 458
Arlington, Virginia 22217

Naval Electronics Laboratory Center — 1 copy
Advanced Software Technology Division
Code 5200
San Diego, California 92152

Mr. E.H. Gleissner — 1 copy
Naval Ship Research and Development
Computation and Mathematics Department
Bethesda, Maryland 20084

Captain Grace M. Hopper, USNR — 1 copy
Naval Data Automation Command, Code 00H
Washington Navy Yard
Washington, D.C. 20374

Dr. Robert Engelmore — 2 copies
Advanced Research Project Agency
Information Processing Techniques
1400 Wilson Boulevard
Arlington, Virginia 22209

Professor Omar Wing — 1 copy
Columbia University in the City of New York
Department of Electrical Engineering and
Computer Science
New York, New York 10027

Office of Naval Research — 1 copy
Assistant Chief for Technology
Code 200
Arlington, Virginia 22217

Computer Systems Management, Inc. — 5 copies
1300 Wilson Boulevard, Suite 102
Arlington, Virginia 22209

Ms. Robin Dillard — 1 copy
Naval Ocean Systems Center
C2 Information Processing Branch (Code 8242)
271 Catalina Boulevard
San Diego, California 92152

Dr. William Woods — 1 copy
BBN
50 Moulton Street
Cambridge, MA 02138

Professor Van Dam — 1 copy
Dept. of Computer Science
Brown University
Providence, RI 02912

Professor Eugene Charniak — 1 copy
Dept. of Computer Science
Brown University
Providence, RI 02912

Professor Robert Wilensky — 1 copy
Univ. of California
Elec. Engr. and Computer Science
Berkeley, CA 94707

Professor Allen Newell — 1 copy
Dept. of Computer Science
Carnegie-Mellon University
Schenley Park
Pittsburgh, PA 15213

Professor David Waltz — 1 copy
Univ. of Ill at Urbana-Champaign
Coordinated Science Lab
Urbana, IL 61801

Professor Patrick Winston — 1 copy
MIT
545 Technology Square
Cambridge, MA 02139

Professor Marvin Minsky — 1 copy
MIT
545 Technology Square
Cambridge, MA 02139

Professor Negroponte — 1 copy
MIT
545 Technology Square
Cambridge, MA 02139

Professor Jerome Feldman — 1 copy
Univ. of Rochester
Dept. of Computer Science
Rochester, NY 14627

Dr. Nils Nilsson — 1 copy
Stanford Research Institute
Menlo Park, CA 94025

Dr. Alan Meyrowitz — 1 copy
Office of Naval Research
Code 437
800 N. Quincy Street
Arlington, VA 22217

Dr. Edward Shortliffe — 1 copy
Stanford University
MYCIN Project TC-117
Stanford Univ. Medical Center
Stanford, CA 94305

Dr. Douglas Lenat — 1 copy
Stanford University
Computer Science Department
Stanford, CA 94305

Dr. M.C. Harrison — 1 copy
Courant Institute Mathematical Science
New York University
New York, NY 10012

Dr. Morgan — 1 copy
University of Pennsylvania
Dept. of Computer Science & Info. Sci.
Philadelphia, PA 19104

Mr. Fred M. Griffee — 1 copy
Technical Advisor C3 Division
Marine Corps Development
and Education Command
Quantico, VA 22134

Explanation: A First Pass

by

Roger C. Schank

Yale University

People explain their actions to others, and to themselves, every day. Why do we need to send and receive these explanations? What is the value of such explanations? After we explain a thing to ourselves or accept the explanation of another, what becomes of it? What is the point of these explanations? What is their role in the learning process? What do they tell us about what it means to be intelligent?

For some people, explanations do not play a big role. Many people are willing to observe events that would disturb others, and attribute these events to *inexplicable circumstances*. They may not even wonder about them at all. What is the difference between people who search for explanations for everything and those who do not require them? What are they doing differently? Is curiosity the major factor with little more significance than the entertainment value of the explanation? Is there some emotional satisfaction derived from knowing why things have happened the way they did, or, is something more significant afoot?

In this paper, I will discuss some of the issues surrounding the problem of explanation. Mostly, I will concentrate on making clear exactly what I believe the problem to be, and what issues need to be addressed in order to solve the problem. I will do this by concentrating on a number of examples that serve to illustrate the issues. I will only suggest an outline of some solutions however. Full solutions await detailed research.

The Role of Reminding

When we first began to study reminding (Schank [1980], Schank [1982]), it seemed clear that explanations held the key to reminding. Whenever two phenomena were related in a reminding experience, we found an explanation in common. Both events, the **remindee** and the **remindand** were analogous in the same way, usually having in common a given expectation failure. The explanation of the anomolous expectation served as the link between them. Thus, in an important way, explanation and learning were already linked. To make sense of, and correct, an expectation failure, explanations were made. These explanations were indices to the events they explained and, as such, could cause remindings. The result of all this was a *corrected*

expectation or in some cases, a reorganized set of expectations. Thus the link went: expectation-failure ---> reminding ---> generalization ---> learning (modifiation of expectation that failed).

My favorite example of this was the ***Steak and the Haircut*** story:

The Steak and the Haircut

X described how his wife would never make his steak as rare as he liked it. When this was told to Y, it reminded him of a time, 30 years earlier, when he tried to get his hair cut in England and the barber just wouldn't cut it as short as he wanted it.

The argument that I made with respect to this reminding ran as follows: The understander of the steak story must be using some kind of knowledge structure in order to get reminded of the haircut situation. To do this, he must have been using a structure that was general enough to cover both stories. This structure must have contained expectations about what actors commonly do in situations such as this. For our purposes here we can call this structure PROVIDE-SERVICE. The assumption here is that in processing the story about the rare steak, the understander would have used a structure like PROVIDE-SERVICE as a source of predictions about the actions that are likely to come next in the story.

Since the predictions contained in a structure such as PROVIDE-SERVICE are about the behavior of the participants in the situation governed by that structure, the understander can be assumed to have predicted here that when someone who has assumed the SERVER role in that structure voluntarily, that he or she will do what he has been asked to do if he can and if what he has been asked is within the domain of the area in which he normally provides his service.

In the steak story, making steak rare is within the range of abilities of the SERVER, yet she has failed to do so. This is an expectation failure. Our thesis was that such failures must be explained. So, the problem for the understander is to explain why the prediction that was made was in error. *Why didn't the server do what she was asked?*

There are many possible avenues of explanation here. The SERVER could be feeling hostile, recalcitrant or whatever. The correct explanation is not important. What does matter is how Y explained things and how that explanation served to remind Y of the experience with the barber. Y must have assumed initially that the SERVER intended to do what the SERVEE wants, and, having found that assumption to be in error in this case, he had to create an explanation that accounted for the behavior of the wife of X.

There are many possible explanations, but Y seems to have used: *SERVER must not believe*

that SERVEE wants what he said he wants, he must want something less extreme. In constructing such an explanation, an index to memory was also created. That is, sometime, thirty yearlf earlier, Y must have explained the haircut story with the same rule. Now that rule has been waiting all this time, to be used as an index to that story if it were ever needed. That is, Y decided (subconsciously we assume) to remember that story as an instance of a correction to an expectation he had.

The premise of our earlier work on reminding was that learning occurs as we gather up failed expectations and correct them. Explanations, we hypothesized, are used as indices to prior experiences that have failed in similar ways. Comparing two such stories can lead to learning.

Now, as we have progressed in our work on reminding and memory organization, it's becoming clearer that we missed the mark somewhat. The above chain is right enough, learning does seem to occur in this way. But, the focus is wrong. Previously we focussed on the value of reminding and the correction of expectation failure, whereas we should have been focussing on the explanations themselves. Here's why:

Intelligent human beings like to understand the people with whom they interact to the extent of knowing what they may do and why they may do it. They seek to understand the institutions that they deal with. They want to know how to treat the rules that these institutions set up and how the institutions will treat them. They also want to know how the physical world behaves. They seek to understand why machines behave the way they do and how physical objects and forces can best be dealt with.

To summarize: people want to understand the world - personally, socially, and physically.

Making Explanations

A great deal of this attempt to explain the world is tied up with the notion of generalization. We don't just seek to know why a given person does what he does, although we may accept an explanation that pertains only to him if that's the best we can do. We also want to know how this new rule that we have just learned can apply to other, similar situations. We seek to generalize the behavior of others in such a way as to create rules that will hold in circumstances other than those we have just encountered. If we are successful at a stock purchase, for example, we wish to know if our success was due to our keen insight, our broker, the day of the week, the industry our stock belongs to, the nature of the market, the weather, or whatever. If we want to replicate successful behavior then we must know what that behavior was. Behavior is so complex that just because the result was successful it doesn't follow that we can easily repeat what we did. We may have done a great many things, most of which were probably irrelevant. (For

example, my uncle, who was a successful football coach, always wore the same brown suit to the games. I assume that he knew in some sense that this suit was not the reason that he was successful, but he replicated everything that he could.) We need to know what aspects of an event are significant and which are relevant with respect to what we can learn from the event for the future.

If we wish to account for failures, then when we do fail, we must explain our failures in such a way as to be able to modify the aspect of our behavior that was in error. Finding just which aspect is most significant can be a serious problem however. We must know how to generalize correctly. Thus, we must come up with an explanation that correctly covers the range of behaviors where it will be most useful. Our explanations must be inclusive and instructive. They must include more behavior than we just saw and they must instruct us on how to behave in future situations of a like kind. Establishing what kinds of situations constitute a like kind is one of the main problems of generalization. It thus is, in some sense, the purpose of explanation.

Not every explanation is instructive or inclusive. Sometimes we explain things to make sure that they are not of interest. This is one reason why the explanation process must be more critically examined than the reminding process. We do not get reminded every time we attempt *to explain something*. Not all explanations are so significant as to cause a reminding.

We do a great deal of explaining without learning a thing of interest. In other words, explanation is going on all the time, it is a much more pervasive phenomon than either reminding or learning and we must examine what starts it, how it is accomplished, when the result is pursued to the extent that it is generalized and causes learning to occur, and when a resultant explanation causes us to drop a line of inquiry as a target of further learning. Most important of course, is to establish when and how it is pursued to the extent that the explainer is satisfied that an anomalous situation is no longer anomalous.

One thing about explanation that is important and different from other processes, is that we know the kind of explanation that we seek before we start the process. With reminding and learning we often are surprised with what comes out; we don't really know where we are going. With explanation, we know one when we see one. So, what suffices as an explanation is of key importance in discussing the nature of explanation. We must know the nature of satisfactory explanations beforehand in order to be able to be sure that we have one. In other words, a sense of the coherency of what we have processed is critical in the explanation process.

The Explanation Process

Roughly, the explanation process involves the following:

1. Find an anomaly
2. Establish the kind of explanation that will make it less anomalous
3. Formulate the explanation pattern that will suffice
4. Explain
5. Take explanation and establish whether:
    a. it makes clear the anomaly but does no more than that
    b. it must be generalized beyond the current case
6. If we must generalize, then find the right item for generalization
7. Write new rule that has just been formulated
8. Find breadth of its application (scoping)
9. Verification (often by reminding)
10. Reorganize at a greater level of generality.

Reminding as Verification

Clear from this list is the role of reminding in the process. If reminding occurs it is one method by which the generalization of an explanation can be justified and through which the new explanation can be used at a high level to reorganize some rules in memory.

As an example of this consider the following:

EXAMPLE 1

> I was walking along the beach in Puerto Rico and noticed signs saying that it is unsafe to swim yet everyone is swimming and it is clearly safe.
>
> I explained this to myself, after seeing a second sign of a different sort, warning about the dangers of walking in a given place, by assuming that the hotel that put up these signs was just trying to cover itself legally in case of an accident.
>
> At this point, that is, after the explanation, I was reminded of signs in Connecticut that say *road legally closed* when the road is in full use. I had previously explained these signs to myself in the same way.

Here we see a classic case of the real role of reminding. First an anomaly is discovered. Next an explanation is concocted. The role of reminding here is as verification. The reminding serves to convince the mind that the explanation that was concocted is reliable. It also gives potential for scoping the generalization that will be formed from the explanation. Here we see that both a state (Connecticut) and an institution (a hotel) can make the same rules for the same reason. Thus our new rule has to be generalized high enough to cover *institutions who could have liability under certain circumstances*. The trick here is to not over-generalize. We learn from

these examples that some signs should be ignored. But which signs and under what circumstances? We want to learn to ignore signs some of the time but not all of the time. Should we ignore stop signs, or signs asking us to register at a hotel? Clearly not. Honing the rule so that it correctly applies is an important part of the explanation process.

The role of explanation-by-example is thus crucial in reminding. People learn better by the use of examples, that much is obvious. The reason that this is so has eluded us, but reminding makes it clear that we construct our own examples to help in learning a new rule in memory. What seems obvious is that the rules we know are grounded in sets of examples.

Finding Anomalies

The next question before us is how explanation works. Why do we choose to explain something and what do we do with the explanation? It seems clear that reminding will be available as verification in only a small proportion of the explanations that we do. In unverified cases (that is, unverified by reminding), we may look for other types of verification such as seeing if our explanation meets certain standards of coherency for explanation. Thus - *the partner of reminding is coherency* or making sense of a new explanation.

What situations need to be explained? People have powerful models of the world. Through these models, which are based on the accumulated set of experiences that a person has had, new experiences are interpreted. When the new experiences that a person perceives fit nicely into the framework of expectations that have been derived from experience, an understander has little problem understanding. However, it is often the case that a new experience is anomolous in some way. It doesn't correspond to what we expect. In that case, we must reevaluate what is going on. We must attempt to explain why we were wrong in our expectations. We must do this or we will fail to grow as a result of our experiences. Learning requires expectation failure and the explanation of expectation failure.

But, expectation failure is not a simple process. When we have only a few expectations and they turn out to be incorrect, finding which one failed is not that complex a process. In the real world however, at any given moment we have a tremendously large number of expectations. In fact, people are constantly questioning themselves and each other, in a quest to find out why someone has done what he has done and what the consequences of that action are likely to be. Thus, in order to find out how we learn, we must find out how we know that we need to learn. In other words, we need to know how we discover anomalies. How do we know that something did not fit?

The premise here is that whenever an action takes place, in order to discover what might be

anomolous about it, we have to have been asking ourselves a set of questions about the nature of that action. In other words, we are constantly, during the course of processing, asking certain questions about that event, in order to fully understand it. Anomalies occur when the answers to one or more of those questions is unknown. It is then that we seek to explain what was going on. It is then that we learn.

To get a handle on this process, we must attempt to sort out the kinds of anomalies that there are. Knowing the kinds of anomalies that there are gives us two advantages. In order for us to find something to be anomalous we must have been unable to answer a question about some circumstance. So, first we must discover the questions that are routinely asked as a part of the understanding process. Second, in finding out what anomalies there are, we also have the basis for the kinds of explanations that are created to take care of those anomalies. Thus we understand what can be learned.

Since we learn from everything, by the reasoning above it follows that everything can be anomalous. But what is *everything*? The **things** we seek are the types of events that there are in the world. For example, we observe the actions of others in the world around us. To find anomalies (or more directly, to understand what they are doing), we ask questions of ourselves about their actions. For actions by individuals, I propose the following set of questions, which are asked in some sense, every time that an action is observed:

1. PATTERNS: Is this an action that this person ordinarily does? Have I seen him do it before? Is it an action that a member of a group that I classify him in ordinarily does? If not then...
2. REFERENCE TO SELF: Is this an action that I would do? If not then...
3. RESULTS: Is this an action that will yield a result that is clearly and directly beneficial to the actor? If not then...
4. PLANS: Is this action part of a plan that I know to be a plan of the actor's. If not then is this an action that is part of an overall plan that I was previously unaware of that will, in the long run be beneficial to the actor? If not then...
5. GOALS: Is this an action that might be determined to be effective in achieving a goal that I know this actor has? If not then is this action helpful in achieving a goal that I did not know he had but might plausibly assume that he might have? If not then...
6. BELIEFS: Is there a belief that I know that the actor holds that explains this action? If not, is there a belief that I can assume he might hold that would explain this action?

In the end, the result of this process is either a new fact, (a plan, goal, or belief that one did not know that a given actor had), or else the action is unexplainable.

The theory is that every time that someone does something, an observer, in his attempt to interpret the action that he is observing, checks to see if that action *makes sense*. But, actions do not make sense absolutely. That is, we cannot determine if actions make sense except by comparing them to some standard. Thus, we must propose a standard.

The standard that we believe to be in use in these circumstances corresponds to three issues that we believe to underlie most observations. These are:

PATTERNS

CONSEQUENCES

REASONS

In other words, we are satisfied, as observers of actions, when the action that we observe fits into a known pattern, has known consequences that we can determine to be beneficial to the actor, or is part of an overall plan or view of the world that we can ascribe to the actor.

To put this another way, when we see that something has done something, we try to find the pattern to which it belongs. Failing that, the consequences that will result become an issue. If those consequences are beneficial to the actor then nothing needs to be explained since there is no anomaly. If those consequences are not obviously beneficial, then we need to find out why the action has been attempted in the first place. This requires ascertaining what goals an actor has, what plans he believes will effect those goals, or what beliefs he has from which a goal may have been generated.

The premise here is not that people are trying to find out whether something is anomalous and needs explaining. In fact, quite the opposite is the case. An understander is trying to determine the place for an action that he observes. His goal is to find a place for the representation of the action within the context of the other representations that he has in memory. To do this, he must find a place in memory that was expecting this new action. Of course he may not find one since not everything in life can be anticipated.

So, an understander asks himself the question, ***what structure in my memory would have been expecting this action had I reason to believe that that structure was active?*** It is at this point that the above issues arise. In other words, we are always asking ourselves why things have happened the way they have, but most of the time we know most of the rudiments of the answer since what we are observing is fairly usual. When things are unusual and must be explained, it is because some of the usual questions have gotten some unusual answers.

What Must be Explained

We will not attempt to provide a complete list of questions that people ask themselves about the world around them. In general, there do seem to be three classes of things that people explain. One is the physical world, one is the social world, and one is the specifics of the individuals that inhabit the world.

The kinds of explanations we expect to get constrain the explanation process by giving us a hint of what to look for. Things are just not anomalous or coherent. We understand that there are limits to our knowledge and we seek to explain what it is that we don't know. The first broad class of things that we shall look at are physical explanations. We don't know all there is to know about how the physical world works. So, we occasionally find anomalies in our daily lives that we seek to explain. Below are some examples that I gathered from students at the Yale AI lab. They are intended to illustrate the types of physical anomalies that people find in their daily lives and the explanations that they conconct and accept that are then added to their memories:

Physical Explanations

EXAMPLE 2 (The Ice Storm)

During the ice storm last night, Suzie and I were in my apartment. Neither of us had realized that the snow had turned to freezing rain. We heard a long series of crackles and whooshing sounds. I said that it sounded like trees falling. During the ice storm of '79 (my first real winter) I was nearly killed by a falling limb, and I suppose I am now sensitized to that noise. Suzie was sure it wasn't trees because there were so many similar noises. When we awoke, it turned out to be dozens of fallen trees.

EXAMPLE 3 (Car Doors)

I've been spending more time in Alex's little Honda these days, and I never seem to be able to close the door completely. Whenever I get in or out of the thing, the dome light stays on to tell me that I haven't managed to close the door tight. Alex, on the other hand, never seems to have this problem. Last week, for the first time, I slammed the door hard enough to close it, but Alex didn't close his all the way. It struck me then that the problem wasn't me, but the second person to close a door can't do it. After pondering a moment, I decided that it must be air pressure - when the car is sealed save one door, the pressure in the car keeps that last door from closing all the way. When another door is opened, the air has another place to go.

EXAMPLE 4 (Snow Patterns)

Up around Kline Biology Tower this morning I noticed the following phenomenon. Along the causeway there are a number of large, round brick columns. Around these columns there was snow, as everywhere else. However, for about 1 foot outward from the base of each column, all the way around, there was no snow. In other words, the wind had cleared away the snow for some reason all the way around the column, even though it was eight inches deep at the edge of the clear ring.

The reason this needs an explanation is that the same phenomenon is not observed at a plain flat

wall. The first explanation we came up with was that the wind hits the column and then blows up, down, and around it. The wind that is blown down blows the snow away from the column. If the wind comes from different directions at different times, the around will be clear all the way around the column.

However, I was unhappy with this explanation, because it does not use the feature of columns which makes them different from walls. If this explanation were correct, one would expect to see clear spaces along the bases of walls, too, but this is not the case. I then came up with another explanation, which we agreed was more likely (although we still don't know if it is correct). Wind hits the column, and because of the circular shape, is more likely to blow to either side than up or down. The wind blowing along the side of the column adds its force to the wind that was already blowing there, giving you stronger winds along the sides of the column. These strong winds clear out the snow for a small area along the sides of the column which are parallel to wind direction. Given shifting winds, we would expect to see the snow cleared all the way around the column.

EXAMPLE 5 (Slipping on Snow)

This morning, as I stepped out the door into the snow, I expected it to be slippery but found instead that the snow was nice and crunchy -- better than average traction. Walking to my bus stop, I became aware that it was a bit late and I had better hustle if I wanted to make the bus. I started to speed up to a jog and went into a skid, almost falling. I wondered what had happened to my nice crunchy traction. I thought that it might be that the weather was changing. This didn't seem likely since it had only been three minutes. I then thought that maybe the snow on this block was different. I have noticed in the past surprising differences in this few-block-area of how early flowers bloom, how soon snow melts, etc. This reminded me of ice I had noticed in that part of route previous day. While most of the way was clear, I had detoured around a bad-looking patch. Then I decided that the snow itself was still high-traction. What was slippery was the ice underneath. I was slipping on old ice, which I had easily avoided when I could see it, but which now looked the same as where there was concrete under the snow.

EXAMPLE 6 (Lightbulbs)

The lightbulb in the hall blew for the second time in a month. Diane said that we must have put in an old bulb. I decided that it was 100 watt bulb in a closed container and that heat blew it. I put in a 60 watter and decided to give more strength to my belief that light bulbs in a closed space cause too much heat to accumulate which causes them to blow out.

What do these examples have in common? They all are attempts by an understander to better understand the world around him. Clearly there is a survival mechanism at work here. We cannot function effectively if we don't understand the world we live in. We need to know about how to deal with winter (3 of the examples) for instance. We are constantly learning about winter (the more experienced we are, the better we survive). So, we make hypotheses. The explanation process is thus critical to survival in the physical world.

From the above, we can glean some of the questions about the physical world that people seem to ask. These are:

From example 2 (The Ice Storm): We seek to identify strange noises. Roughly the question we ask is: What was that strange noise? What caused it? Is there danger to me?

From Example 3 (Car Doors): We seek to learn how to make the objects around us work correctly. In this example, the question was how to get a car door to close. The general question is: Why won't this object do what it is supposed to do? The kinds of answers we seek to this type of question are either functionally-specific (that is, facts gleaned about this particular object) or general physics (such as the principles of air pressure that the teller of this story concluded were active in this case.)

From Example 4 (Snow Patterns): In observing the world around us we sometimes wonder why things turn out the way they do. Some people concern themselves with the problems of general physics more than others. This is an example of the continuing attempt of a person to determine the physics of the world.

From Example 5 (Slipping on Snow): As we saw with the strange noise example, people are concerned for their physical well being (naturally enough). One problem is avoiding dangers the next time around. Roughly, the question is, how can make sure that I don't do a second time what caused me pain the first time? Again explanations can be specific or general. A simple explanation in this example might be that it is always icy in one spot so that spot should be avoided. But obviously there is no general utility in such an explanation. So, we seek to identify how a danger can be hidden from view. The question we ask therefore is what dangerous situation might we be able to avoid by understanding how and why a danger can be hidden from view?

From Example 6 (Lightbulbs): Here we want to preserve the physical status of objects. We also wish to prevent dangerous circumstances from occurring. The questions arising here therefore are: How can I change the circumstances that have been endangering an object I value such that it will not be harmed? Also, how can I prevent danger by attempting to better understand the physics of the world in general?

Physical Questions

To summarize then, the following is a list of representative types of things that people care about with respect to the physical world around them. People feel a need to explain any event for which there is no straightforward answer to these questions. The questions are:

1. Is there something that caused a strange noise that might be of danger?
2. Why doesn't an object object do what it is supposed to do?
3. What general principles of physics explain why things are happening the way they are?
4. How can we avoid a previously encountered dangerous situation?
5. What might cause a dangerous situation to be hidden from view?

6. How can I change the circumstances that have been endangering an object I value such that it will not be harmed?
7. How can I prevent danger in general by attempting to better understand the physics of the world?

Social Explanations

It is also critical to deal with the society in which we live. We want to know why the institutions that we deal with behave in the way they do. In this way we can better interact with them. Some examples:

EXAMPLE 7 (Hairdressers and Credit Cards)

Diane was trying to figure out why hairdressers won't take credit cards. She thought that maybe they had a poor clientele but realized it was also true in Westport. She never found an answer.

EXAMPLE 8 (IBM Policies)

The conflict:

- IBM sells foreign-language versions of their word-processing software in foreign countries, but refuses to sell them in the US.
- There are people in the US who would buy the foreign-language programs.

Explanation: Since IBM has a vast and sophisticated marketing department, they must have determined that it's not cost-effective to distribute the foreign-language software in the US.

EXAMPLE 9 (Reporter in Lebanon)

This morning, watching the Reuters newswire on TV, one of the headlines at the beginning read *American TV network reporter feared kidnapped in Beirut*. Immediately, Jerry Levin, CNN's reporter in Beirut, came to mind, not as a conscious prediction, but just sort of idly. When I first noticed that he had been posted as Beirut bureau chief for CNN, I remember thinking that it was a bit risky for a Jew to take that assignment. Sure enough, it turns out that it's Jerry Levin who is missing and believed kidnapped.

EXAMPLE 10 (Campaign Predictions)

I noticed with Hart all the news people yesterday were saying that *that's what he's been predicting for two years* as though he had such a sage understanding in advance of how the campaign was going to work. That's how they explain the situation, or at least provide background.

I got very annoyed as different news shows repeated this and I wondered why NONE of them were swift enough to look at other predictions made by other candidates of how THEY would do and blow the whole notion away. I explain this anomaly by positing they just want something to say to fill up air time and are too dumb to think about what is really going on.

EXAMPLE 11 (Yale and Tenure)

At dinner last night, we were talking about the woman who was denied tenure at Yale (History Dept.) and sued the university. Yale settled out of court.

- The woman claimed that she was denied tenure because she is female, and that they gave

her position to someone less qualified.

- For academic reasons, the woman did not deserve tenure (opinion of the history faculty). Her position was given to another woman. Courts cannot decide academic qualifications. One would have thought that Yale would have fought it out in court, defending priciples. Why didn't they?

Explanations:

1. It's cheaper to settle out of court than to let the case drag through the courts
2. Yale settled out of court precisely BECAUSE courts can't decide academic qualifications, so Yale gets to say (implicitly) that the woman was a poor scholar, although they chose to pay her off.

We try hard to understand and predict the actions of the social institutions that make up the world in which we live. Here again there are questions that we ask in our attempt to deal with the events that we witness. When these questions give satisfactory answers, we fail to notice that we ever asked them. However, we are forced to answer these questions by seeking an explanation when the answering of them is in some way complicated. This complication usually arises when no pre-formed answer exists.

Here then are some of these questions, as derived from the above examples:

From example 7 (Hairdressers and Credit Cards): When the goal of an institution is clear (businesses want to make money), any impediment to the achievement of that goal needs explanation. Also, the rough form of a solution to any quandry about a business can usually be determined within the given parameter that, whatever the decision by a business, it probably was made in the attempt to make money. Therefore the operating question whenever a business is the institution involved is: How can the decision that a business made be seen as a way of making more money in the long run? We will return to this later.

From example 8 (IBM Policies): Given what we said above, whatever the policy of IBM, the real question is how that policy makes them more money. This was in fact the explanation given.

From example 9 (Reporter in Lebanon): In an attempt to understand an event we are wont to predict future events that may derive from an event we observe. Thus, one question we ask all the time is: What will the consequences of an action be? Are those consequences in conflict with known goals? In this case the predicted danger to the reporter was predicted by the actions of a well-known political institution (the PLO?). The question that the writer of this story was asking was how anyone could be so stupid, but he was not really expecting an answer since human stupidity is a common enough commodity.

From example 10 (Campaign Predictions): There is a general question of why networks behave as they do. Here again, since institutions have known goals pursuant to their purpose as institutions, we can assume that the networks perform according to those roles. Thus, we assume that the networks, in general, wish to inform about the news, and make money in so doing. Given a certain amount of air time to fill up, and a lack of available expertise at all times of the day, they frequently must *fill up air time*. The sense of what they say may be held as being of only secondary importance. Thus in essence the question here is: What is the order of priority for an institution given a complex set of (possibly conflicting) goals?

From example 11 (Yale and Tenure): Yale is, of course, an institution whose goals are known but are different from most other private instititions. On the assumption that Yale's goals are different enough that it must defend its principles every now and then, this observer wondered why Yale hadn't acted in usual fashion. His explanation was based on known principles of Yale's behavior. The question he asked was: Why would an institution that has a set of well-established principles not choose to defend them?

Social Questions

To summarize again, we have derived the following general issues that people track in dealing with social institutions:

1. What is the expected behavior of a social institution?
2. When is it in violation of that behavior?
3. What will the consequences of an action of an individual be when it interrelates with those of an institution?
4. Are those consequences of the action of an institution in conflict with its known goals?
5. How can a decision that a business makes be seen as a way of making more money in the long run?
6. What is the order of priority for an institution given a complex set of (possibly conflicting) goals?
7. Under what circumstances do instititions abandon their established patterns of behavior?

Patterns of Behavior

Most explanations however deal with our attempts to understand the specifics of the world around us. These specifics include the objects we deal with and the people we deal with. We expect certain patterns of behavior and spend time trying to find out which people belong to which groups in their behavior patterns, and what those patterns are: Here are some examples:

EXAMPLE 12 (Holding the Pillow):

I was spending a quiet evening with Suzie, who is in the midst of an internship in midwifery. She had had a difficult day (delivered her first dead baby) and was finally relaxing a little. We were sitting on the bed and she was hugging a pillow to her breast. It was not a focus of her attention, she was just holding it there as we talked. This reminded me of the fact that I had seen someone holding a pillow like that before; I didn't remember the particular scene, but I did remember that the other person had also been a woman. I asked her why women do that -- she was surprised, and said she hadn't even noticed she was holding it, but that it somehow made her more comfortable. The feeling was quite palpable, and changed when the pillow was held differently. My explanation, which she agreed with, was that it affected the level of some hormone associated with nursing; that would make holding things in that position feel good. The pillow is roughly baby sized, and the phenomenon only worked in a narrow range of positions. There should be some sort of biological mechanism for making nursing attractive. Suzie's might have been activated by her work (or always present in women of that age) and she could have been sensitized to the feeling by her stressed out, emotional condition. We had no other explanation of why that sensation should be so strong, or why neither of us could think of a man in a similar position.

EXAMPLE 13 (Shoeless Students)

Yesterday I walked into a graduate student office and saw David sitting down reading the paper with no shoes on. I chalked it up to idiosyncratic behavior. Ten minutes later, I was walking down the hall and saw Jonathan walking toward me barefoot. At that point I decided that there must an explanation. I realized that it had been raining very hard that day and that David and Jonathan must have gotten their shoes and socks soaked. They had then taken them off to let them dry. Jonathan confirmed my hypothesis.

EXAMPLE 14 (New Hampshire Primaries)

Political postmorteming over the Gary Hart win in New Hampshire reminded me of McCarthy in 1968. In February in New Hampshire, not much is happening other than waiting for the sap to rise in the maples, so voters are receptive to the energy and youth of young campaign workers. Similarly for the weather. It was snowing, so the candidate with the most 4-wheel drive owning supporters was likely to win.

EXAMPLE 15 (The Parking Lot Gate)

Yesterday, walking to my car, I approached the automatic parking lot gate to my parking lot. The gate is the type where you insert a plastic card to get in, and that automatically opens when you approach the gate in your car to get out.

At a distance of about 50 feet I noticed that the gate rose about half way up and went back down again. I also noticed that there were no cars moving in the vicinity. The spontaneous spasm of the gate was highly unusual. I have approached this same gate on foot countless times and this never happened before.

I tried to explain why the gate behaved in this way. The first thing I did was to check again that I did not miss a car that was approaching the gate. I saw no car, but I did see a government police car parked in the distance in a place where cars do not usually park. I then wondered if the police car had something to do with the gate's behavior. Did they have some device that triggered the gate? Were they watching the gate? I decided that this was unlikely and paranoid.

I then wondered if I had not in some way triggered the gate in my approaching it. I was reminded of the times that I passed through airport metal detector gates and triggered the alarm due to something I was carrying. This seemed to make more sense so I pursued it further. was the parking gate triggered by some sort of metal detector? This seemed plausible, but cars have a

great deal of metal in them and I do not. Was I carrying something with metal in it? Yes, but no more than I usually carry and the gate had never mysteriously moved before when I passed it. Furthermore if a car were the distance I was from the gate it would not have triggered the gate. I rejected the metal detection explanation as well.

I decided that for the purposes of the moment the gate was acting unusually because it might be broken. I hoped that the gate would work when I tried to get my car out of the lot. Fortunately, the gate did work.

I told my wife the story of the gate and she came up with two explanations that I hadn't thought of. One was that someone left their card in the device that reads the cards and this was causing the aberant behavior. The second was that it had been raining heavily that day and perhaps the rain caused some electrical problem in the gate.

EXAMPLE 16 (Barking Dog)

Walking home, I was deep in thought when I heard the sound of a dog barking fiercely and saw it coming at me. Without stopping to think, I slipped my hands into the pockets of my coat and, without breaking my stride, made eye contact with the dog and asked: *Just what do you think you're doing? Hunh? HUNH?! Just WHAT do you think you're doing?!* Startled by the sound of my own voice, I suddenly felt quite foolish. What was I doing speaking out loud (asking silly questions, no less!) to strange, hostile dogs in public places? Discretely glancing around, I was relieved not to detect any witnesses. Then I realized that that *it had worked* -- the dog had shut up, backed out of my way, and was looking totally cowed and confused.

What was it I had just done that deflected the dog crisis?
Why did it work?
How did I know to do it, given I was still first realizing what the threat situation WAS?

1. My initial explanation was that I had CONFUSED the dog.
2. My second explanation was that I HADN'T ACTED FRIGHTENED. A rule I had been taught as a child about dealing with hostile dogs is: don't let them know you're afraid, they'll only bite you if they think you're afraid of them. I had done just what I was supposed to do, just like I had been told. And it worked, just like they said it would.
3. I had refused to go along with the dog's suggestion that I play the frightened-victim role, and instead had selected the dominating-human role. That forced HIM into the *subordinate-dog role, eliminating his aggressive options.* (I had had a problem with a bothersome dog when I was a kid, and out of desperation to protect my own dog, whom I had already scooped up and was trying to hold out of reach, frustratedly told it to SIT! And it suddenly became all meek and apologetic -- and sat! Warning it to *Stay, staaay. Bad dog! sit! Sit! Good dog! Now stay!* we made our exit. What I had learned from this is: dogs with collars are already trained to obey SOMEone, already have the concept of human authority.)

EXAMPLE 17 (Simultaneous Primaries)

Jim questioned why the primaries dragged on so long, complaining about how they dominated the news, making it harder to find out what ELSE was happening in the world. He advocated a much shorter primary season, suggesting that best of all would be if they were all the same day. My explanation: The primaries issue skeleton is under the control of each political party. It's in the interest of the political parties to drag this on as long as possible, because it is a way of getting free press coverage. The candidates don't have to PAY for advertisement, just hold yet another debate, and the media will compete over who can cover it most.

Here again are some questions that may have been operating that would produce the kind of explanation behavior exhibited in the above examples:

From example 12 (Holding the Pillow): Men often ask about why women behave the way they do and women often ask about why men do what they do. The same is true of Catholics and Protestants, the English and the French, and any other groups that you might want to put in opposition. What we have in this example is an attempt at a serious explanation, but the basic question is: why does someone from the group that I am not in, do what no one from my group would do?

From example 13 (Shoeless Students): Often in our attempt to group people into behavior-prediction groups, we can come up with better explanations. The two people concerned were computer-types of course, so it is easy to say that computer-types are likely to go shoeless in an office (which they are). But finding a cause helps. The question here then has to do with the individuals in question, namely: Why would someone do something odd when I have not noticed that odd behavior? And: why is an odd thing happening more than once (where chance seems a poor explanation)?

From example 14 (New Hampshire Primaries): The pattern of behavior of groups is often interesting where an outcome is known but the group is not. Thus, in voting, when there is an upset, we often hear explanations based on how various groups behaved and why. In this case the question is: Why was a particular outcome beneficial to a given group. Also: what random factors might have caused a given group not to try as hard or to not be able to perform as well as one might have expected?

From example 15 (The Parking Lot Gate): Not only people exhibit patterns of behavior. Objects do as well. When an object fails to perform as expected we seek to find out why. The question is: what could be causing something that ordinarily happens in a certain way to not happen that way, especially when the object is presumed to have no will of its own?

From example 16 (The Barking Dog): All this is true of animals as well as people and machines. We expect one dog to be like another, but we have, at best, a poor model of why animals do what they do. Occasionally it can become very important to know what an animal will do however. The questions are: What can I do to control a wild animal? Why does an animal behave the way it does? What patterns of behavior are operating for given classes of animals?

From example 17 (Simultaneous Primaries): Often we find that an annoyance keeps occurring

regardless of how absurd it may seem. Often the explanation is that circumstances have combined to make it almost impossible to prevent. In other words, there is no rational person in charge who can call a halt to the absurdities. (I am reminded of gas price-wars here, an oddity that has faded into the past.) Questions that bring such circumstances to the fore are: why does a person or institution seem to be doing what is not in their best interests? Or: how are a set of circumstances able to control a situation? Or, closer to the actual case in example 17, who benefits from a circumstance that seems to annoy almost everyone?

Patterns of Behavior Questions

To summarize, we have derived the following general issues that people track in dealing with patterns of behavior. The list given below is certainly not exhaustive. There are many more questions about patterns of behavior that people ask that produce explanations. On the other hand, this list will converge. There are not as many questions as there are situations to be explained. The stories that we have dealt with so far are representative of the type of phenomena that must be explained.

1. What behavior can be predicted by knowing that an individual belongs to a given group?
2. Why does a given group behave in the way it does?
3. Why does another group fail to behave that way?
4. What causes new odd behavior to appear?
5. When odd behavior comes in pairs what is the common explanation?
6. How can given groups effect outcomes beneficial to that behavior?
7. Why was a particular outcome beneficial to a given group?
8. What factors might have caused a given group not be able to effect its desires?
9. Why does a given object stop doing what it has previously done?
10. Why do animals behave the way they do?
11. How can animals be controlled?
12. What patterns of behavior are operating for given classes of animals?
13. Why does a person or institution seem to be doing what is not in their best interests?
14. How are a set of circumstances able to control a situation?

Types of Explanations

We are concerned then with what to explain. The above classes and their associated questions indicate the type of things that need explaining, and what constitutes an explanation. The following is a broad class of explanations that tend to satisfy us when we hear them. That is, once given one of these explanations, we tend to accept them.

1. Excuses

When someone fails to do what we wanted them to do, we complain. They explain by giving a fact that contributed in some way to their behavior. This fact is an excuse. Further, we know that it will be an excuse before it is said. What makes it an excuse is that an excuse is needed. We don't have to analyze it to see if it is an excuse. We do analyze it to see if its a VALID excuse. Invalid excuses are likely to be collected as anomalies for reminding purposes.

2. Alternative Beliefs

When someone does something that we had no reason to expect they would do, we try to find out why by trying to simulate their reasoning to see what they might have been thinking. This type of explanation tells us to modify the belief that we thought an individual held. All predicted behavior stems from what we believe about what another person beleives. We are therefore constantly constructing models of why people do what they do, and modifying them. Thus, since we know that we do not know everything that is another person's belief system, we accept explanations that inform us with respect to our incomplete knowledge.

3. Laws of Physics

Since we do not know all the physical laws of the universe, we often find ourselves speculating as to why something physical has happened. We change our rules to correspond with experience. Hearing a new law of physics therefore can often satisfy us as an explanation. We expect that the physical rules we know will change over time, not because the world is changing, but because our knowledge of the world is incomplete.

4. Institutionalized Rules

When we know that someone is playing according to externally defined rules, we can look for explanations of his behavior in those rules. We assume that we don't know exactly what all those rules are, therefore we constantly update them. This kind of explanation is analogous to #3 except that the rules are defined by people not the physical world. Here again, our understanding of these rules is likely to be incomplete.

5. Rules of Thumb

There are a set of tricks for living that get people where they want to go. *Ask for advice,* or *never date the boss' daughter* are examples. We try to pick these up as we go. Thus, in success or failure explanations we try to validate or add to our rules of thumb. Here we are not dealing with expectations but actual successes and failures. We want to know why Joe always does well or why Sam always fails. We look for a rule of thumb explanation when we want a simple

heuristic rather than an operating principle. It is not necessary to know why a given rules works in order to use it. What is learned here modifies already extant rules.

6. New Facts

We want to learn about things in the world as well as people. We need to know about cats, dogs, cars, computers etc. Again, we learn by explanation. Here, classification is important. When we find a new rule, we try to get it in the right class. When a machine catches fire, our explanation deals with why. One can look at the physics in such cases, but explanations about new-fangled machinery seem to many people to be quite adequate. By putting an explanation in a class of objects, we make it more usable.

7. Appearances

People do things because of how they imagine it will make them look. They buy cars, dress, even marry because it evokes an image of them in their own or others minds that they would like to maintain. Since people do things for this reason, we can often explain their actions in these terms as well. The learning that takes place in these instances relates only to the objects in the universe that we are particularly concerned with. We learn about particular actors in our memories and are thus able to make better predictions about what they will do next time.

8. Plans

We can explain the actions of others by understanding where the particular action that we do not understand fits within a broader plan. Saying that an action is a step on a coherent plan towards a goal, explains that action. Two kinds of learning take place using this type of explanation. We can learn about how a plan is constructed in general. Also we can learn about the kinds of plans that a particular individual is likely to use in a given circumstance.

9. Goals

An action can be explained by connecting it to the goal it was intended to achieve. Knowing what someone wants is an important part of understanding. Accordingly, explaining how an action makes sense in terms of what someone wants is the plan explanation, explaining how it tells us what he wants is the goal explanation.

10. Role Themes

Knowing what role theme a person is acting under tells us a lot about why he is doing what he is doing, and thus pointing out a role theme can be an explanation. That is, saying that a man is

a doctor and that a particular type of behavior is what doctors do, will serve as an explanation is some cases.

11. Scripts

Since scripts are fossilized plans, script explanation are just simpler versions of what we have for plans.

12. Delta Agency

Doing something for someone else does not require a coherent plan of action. It merely requires that one believe that someone who you want to please wants something. It thus explains an action to say that the real explanation is to be found in someone else's plan.

13. Lack of Alternative Plan

Sometimes people do things because they couldn't think of a better action to do. Usually there is a goal in mind, but the achievement of it may not come easily from the action that was implemented. This is a kind of *explain it away* explanation. It doesn't really work in terms of coherency but sometimes it is the best we can do.

14. Mystical Laws

Not everyone shares the same belief system. Those who are religious or mystical may well believe things that are deemed inappropriate by others. Understanding what they believe is part of the explanation process as we saw in #2. The difference between #2 and #14 is mostly whether or not one agrees with the plausibility of the explanation. The major difference is whether this is part of a pattern of belief that would allow one to predict other beliefs that are also likely to be held by this individual.

To see how these classes might be used, consider a situation where a professor asks his advisee why he has not been working on his thesis lately. It is possible, using the classes of explanation given above to concoct many different types of explanations. Each of these explanations would serve to cause his professor to update his model of this student in some way, that is, in a small sense, to learn something:

WHY AREN'T YOU WORKING ON YOUR THESIS?

1. Excuses - Because I had to celebrate my wife's birthday last night.
2. Alternative Beliefs - Because I don't think it needs work.

3. Laws of Physics - It is so hot that the paper keeps melting.
4. Institutionalized Rules - Yale doesn't require one any more.
5. Rules of Thumb - I have discovered that not writing it gets your professor to the point where he will sign anything.
6. New Facts - It's finished.
7. Appearances - I didn't want to appear stupid in front of the other students.
8. Plans - Not writing it every other week is the best way to stay sane while writing.
9. Goals - Oh you thought I wanted a PhD!
10. Role Themes - I've quit school.
11. Scripts - I just don't do that sort of thing.
12. Delta Agency - My wife is writing it for me.
13. Lack of Alternative Plan - The system was down.
14. Mystical Laws - Lama Dama says that that which is not approached direclty is first finished.

Knowing When an Explanation is Right

One critical question then is this: After finding out that something needs to be explained, how do we know that the explanation that we have either received from another, or concocted for ourselves is valid? Clearly one depends upon the other. If we have the ability to recognize a good explanation when we hear it, that is, if we know when we are satisfied, then the knowledge that we have used in that case can be relied upon in constructing our own explanations.

To get a better idea of what I am referring to here, we shall consider some more explanation stories, this time without the explanations. (You can peek at the end for them if you cannot wait.):

EXAMPLE 18

In a store in Vermont, there is a sign saying that they will close tomorrow for inventory. This seems anomalous as it is the middle of their busy season.

EXAMPLE 19

At a hotel they ask my name as well as my room number when I request a wake-up call. Seems unusual.

EXAMPLE 20

An organization has made an appointment for me that they are apologetic about. It turns out that the man is not with that organization. Why did they make the appointment and why are they apologetic?

EXAMPLE 21

FROM The New York Times:

After dashing down the long flight of stairs to the subway, a woman just missed her train and was exasperated. A guard informed her that she shouldn't worry as he felt a local coming soon.

EXAMPLE 22

FROM The Washington Times:

Whenever Dan Rather does a broadcast from a studio other than the one in New York City, his gray hairs seem to disappear.

Explanation Rules

Now let us consider the explanation process by examining each of these examples in turn.

The explanation process focusses on the available data. In 18, the anomaly is that businesses want to make money and this one seems to be counterproductive. Now the question is: what kind of explanation will make it all come out feeling satisfying, that is, feeling as if something (though not necessarily something cosmic) had been learned?

The available data here is sparse. But, suppose I told you that this store was part of a national chain. That information helps because it changes one of the available data items. One data item was the maker of this bizarre decision. Now we find that the maker was someone other than who we thought it was. The decision is still odd, but now we believe that other factors may have entered into his decision process.

The rule then is what I shall call the INCOMPLETE INFORMATION RULE. If something that needs to explained can be seen as being different from what was originally assumed to be the case, then maybe that different thing can be more easily explained. Or perhaps it needs no explanation at all. So, one step in trying to make an explanation is to try and change the variables, one at a time, to see if some hypothetical event comes out that is easily explainable.

The actual explanation of 18 was that the store was part of a national chain that was headquarted in Ohio where it was most certainly not busy season. All the stores do inventory at the same time. Why this had to be the case I never found out. I created my own explanation that it had to do with taxes. Thus the basic principle behind the anomaly was unchanged. This business did indeed care about business but this was how it had to cope. Nothing crucial is learned when the explanation simply involves picking up uninteresting new facts.

Now let us consider example 19.

One possibility here is that the hotel asked my name simply as a way of making sure that they did not wake up the wrong person. What we have here is a script violation however. That is, when a script is violated, it is usually for a reason. (Although one explanation might be that their

script was different than mine.) Scripts tend to be rather sacrosanct. People do things in a given way because they have always done them that way.

In any case I pondered why they had done that, but as it wasn't that fascinating a problem I soon forgot about it.

Explanation of 19: I assumed that it was just a check to make sure I was the right person. Later when my wife told me she had called and they said I hadn't yet registered, I realized they probably had no record of anyone being in that room.

Here we have a case of a new fact consituting an explanation of an event that had been almost forgotten. Obviously it had not been completely forgotten since I was capable of retrieving the relevant information and connecting it up.

The question here is how I knew that this new information constituted an explanation of the early minor mystery. One method of explanation is COORDINATION OF ANOMALIES. Sometimes when two events occur that both seem anomalous, we assume that they are related, without any hard evidence. These two events were certainly candidates for application of that rule. But, it was also important that the EXPLANATION CLASS be correct. For a script to be violated, it is usually the case that some other aspect of the script had to be violated first. In other words, there isn't a free for all on airplane seating unless some problem occurred that causes the airplane seating script to go awry. The same is true of hotels. It is clear then that the explanation type that I sought was from a SCRIPT VIOLATION RULE, which says to look for an explanation for a script violation in some other preceding condition in the script. In other words, something must have fouled up somewhere, and I was looking out for it.

Now consider 20 (the organation"s appointment example):

Why would this organization be apologetic to me? The first part of this reasoning chain is simple: they must have done something bad to me. That is why people apologize. And, why would this organization have done something bad to me, since my visit there was entirely one of good will? This is also easy to answer: someone must have made them do it.

Thus, here we have two rules of explanation. The first is what I shall call the STATE-EVENT CORRESPONDENCE rule. When there is an extant state (apologetic in this case), look for an event that matches to it (something to apologize for). This an obvious and simple rule. For example, if the dog is cowering in the corner he must have done something bad and his owner will look around for what.

The second rule is one of a category of rules that relate to people's behavior. A very common rule of this sort is QUID PRO QUO. If you want to know why someone has done something, one possibility is that he has done it as what he perceives as an equal response to something you did earlier to him. That rule might be considered and rejected in this case. The rule that wins is the DELTA AGENCY rule, also known as SOMEONE MADE ME DO IT. If you cannot find an equivalent action that caused the event that needs explaining, then one choice is to find an actor who made the other actor do what he did. In a big organization, this might be someone further up in the hierarchy.

With all that having been said, it should now be clear that the explanation process is set up by these rules in this case. Thus, I was set up to look for an event that would be unpleasant for me. Also, I was set up to look for a reason for that event that had something to do with the fact that this man had some power over the people I was meeting.

Explanation of 20: Both turned out to be true. This man was someone who clearly knew very little about what he was talking about and was quite annoying. He turned out to be a personal friend of the head of the organization.

Now consider example 21. Here we have an example of a phenomenon that seems quite out of the ordinary. We need an explanation because most people do not have the belief that you can feel a train coming, particularly a given type of train. What will satisfy us as an explanation?

Earlier, we listed some questions that we claimed are ordinarily asked of the actions of individuals and that, among other things, help us find anomalous behavior. In this case, we have an instance of a belief-based anomaly. That is, clearly this guard has a belief that we do not share. The explanation is either that there is a fact missing from our belief systems or that his belief system is peculiar in some way. Thus, the explantion we seek is either a fact about trains that makes them feelable or a fact about this person that enables him to feel things that others cannot. Failing that, we seek a fact about this person that makes it clear exactly what kind of craziness he has. (However, given that this story appeared in a human interest section of the New York Times, it seems obvious that this latter type of explanation will not be what is forthcoming.)

Explanation of 21: The guard stated that *You can always tell which train is coming by the strength of the breeze down the platform. The local gives off a weak breeze, the express a strong one.*

As we stated above, a new fact will always serve as an explanation. However, in this case the

new fact is not enough. We are forced to make sure that this new fact makes sense. We must ask ourselves why a train would give off a breeze. Answering this requires knowing something about the effects of objects that are goinging through enclosed spaces and the effects of the variant speed of those objects. In other words, an understander, in order to believe the statement of the guard would have to know, or be able to figure out, that expresses travel faster than locals in the New York Subway System.

Now let's consider example 22: What can cause a man's gray hairs to disappear? The simplest kind of explanation is done by use of the CAUSE AND EFFECT rule. We have, in our repetoire of explanations, a set of standard causes and effects. We know that if a tree has fallen, then some form of extreme weather is a likely cause. We know that if a small child has a bruised knee, then he is likely to have fallen down, and so on. These rules are the basis of the explanation process in that they allow us to not have to try to explain everything. Their role therefore is very much analogous to the role of scripts in language processing.

What is unusual about this example is that the ordinary CAUSE AND EFFECT rule is obviated by the fact that whatever procedure Rather may have used to get rid of his gray hairs was unlikey to be applied only when he was out of New York. We are forced to imagine that he cannot get any Grecian Formula in New York City and whatever supply he does have magically evaporates on his entry into New York.

But, the fact that we can postulate such a thing, well illustrates the nature of the explanation process. First we try CAUSE AND EFFECT rules. Next, we try to ACCOMODATE those rules within the confines of what we know about the real world. Thus we are foced to ask ourselves, why, if Rather dyes his hair, he cannot do so in New York. We attempt to find a fact of the world that would make this the case. If we cannot find one, we attempt to invent one. Failing that, we go back and look for other versions of reality.

Thus, th next thing we do is DISTORT reality. By this I mean that we attempt to detrmine which part of the total picture we have been looking at was in error. In other words, we try to change one of the conditions we had to explain as a way of explaining a situation. This is what I call EXPLANATION BY ERROR CORRECTION. That is, there was an error in the situation to be explained. Rather's hair was not different it only appeared different.

Explanation of 22: The lighting isn't as good in other studios. Studio ceilings outside New York are so high that the overhead lights don't catch the gray hairs.

Conclusion

It should be fairly clear, after considering the issues I have discussed here, that the ability to construct explanations is a critical part of our intelligence. We simply cannot learn with the ability to explain. The reason for this is simple. Learning involves hypotheses and hypotheses can be, and often are, wrong. We must be able, as intelligent processors to construct hyotheses about the world. We can operate under these hypotheses for long periods of time, but we must be flexible enough to abandon them when they are in error. Reminding is a way of keeping availble possible errant hypotheses. Explanation is how we correct errant hyotheses.

The message is clear then: THERE IS NO LEARNING WITHOUT EXPLANATION. Consequently there really cannot be an AI program that can be considered to really embody intelligence unless it has the ability to explain why the world that it has to deal with functions the way it does, or, more importantly, failed to function the way it expected in any given instance.

I close with two examples of explanation that I rather like that illustrate something about how people use explanations to teach themselves.

The first is from a book by I.B. Singer entitled **Lost in America**:

> I had been raised to believe that a man with brass buttons, an insignia on his cap knew little compassion, particularly when his victim was a Jew. But Americans and Canadians were different. Why? Did it bear on the fact that Americans and Canadians were richer? Was it the upbringing? Were Anglo-Saxons by nature more inclined to be understanding of another person's dilemma than Slavs or Germans? I was by then mature enough not to seek reasons for the conduct of individuals or even of groups.

The second is from the **Metropolitan Diary** section of the Mar 14, 1984 New York Times, written by Marge Runnion;

> I was sitting with my son and a friend when, amid the snowflakes, a small Oriental rug floated gently down to the sidewalk. We watched for nearly an hour as it became frosted with snow, pondering practical explanations. But the only acceptable theory was simply that it was a magical carpet. My son, whose apartment is furnished in Curbside Modern, was convinced that a greater power had sent the carpet to him.... Everyone says that Manhattan is magical in a snowstorm. Now we know its true.

References

Schank, R. C. "Language and Memory". *COGNITIVE SCIENCE*, Vol 4 no. 3, 243-284, 1980.

Schank, R. C. "Dynamic Memory: A Theory of Learning in Computers and People." *Cambridge University Press*, Cambridge, England, 1982.

END

FILMED

6-85

DTIC