EPISTEMOLOGY AND INFORMATION

Fred Dretske

Epistemology is the study of knowledge--its nature, sources, limits, and forms. Since perception is an important source of knowledge, memory a common way of storing and retrieving knowledge, and reasoning and inference effective methods for extending knowledge, epistemology embraces many of the topics comprised in cognitive science. It is, in fact, a philosopher's way of doing cognitive science.

Information, as commonly understood, as the layperson understands it, is an epistemologically important commodity. It is important because it is necessary for knowledge. Without it one remains ignorant. It is the sort of thing we associate with instruction, news, intelligence, and learning. It is what teachers dispense, what we (hope to) find in books and documents, what measuring instruments provide, what airline and train schedules contain, what spies are used to ferret out, what (in time of war) people are tortured to divulge, and what (we hope) to get by tuning in to the evening news.

It is this connection between knowledge and information, as both are commonly understood, that has encouraged philosophers to use mathematically precise codifications of information to formulate more refined theories of knowledge. If information is really what it takes to know, then it seems reasonable to expect that a more precise account of information will yield a scientifically more creditable theory of knowledge. Maybe--or so we may hope--communication engineers can help philosophers with questions raised by Descartes and Kant. That is one of the motives behind information-based theories of knowledge.

1. Necessary Clarifications: Meaning, Truth, and Information.

As the name suggests, information booths are supposed to dispense information. The ones in airports and train stations are supposed to provide answers to questions about when planes and trains arrive and depart. But not just <u>any</u> answers. <u>True</u> answers. They are not there to entertain patrons with meaningful sentences on the general topic of trains, planes, and time. Meaning is fine. You can't have truth without it. False statements, though, are as meaningful as true statements. They are not, however, what information booths have the function of providing. Their purpose is to dispense truths, and that is because information, unlike meaning, has to be true. If nothing you are told about the trains is true, you haven't been given information about the trains. At best, you have been given misinformation, and misinformation is not a kind of information anymore than decoy ducks are a kind of duck. If nothing you are told is true, you may leave an information booth with a lot of false beliefs, but you won't leave with knowledge. You won't leave with knowledge because you haven't been given what you need to know: information.

So if in formulating a theory of information we respect ordinary intuitions about what information is—and why else would one call it a theory <u>of</u> <u>information</u>?--we must carefully distinguish meaning, something that need not be true, from information which must be true. There are, to be sure, special uses of the term "information"— computer science is a case in point—in which truth seems to be irrelevant. Almost anything that can be put into the memory of a computer, anything that can be entered into a "data" base, is counted as information. If it isn't correct, then it is misinformation or false information. But, according to this usage, it is still information. Computers, after all, can't distinguish

between "Paris is the capital of France" and "Paris is the capital of Italy." Both "facts", if fed into a computer, will be stored, retrieved, and used in exactly the same way. So if true sentences count as information, so should false ones. For computational purposes they are indistinguishable.

This approach to information--an approach that is, I believe, widespread in the information sciences—blithely skates over absolutely fundamental distinctions between truth and falsity, between meaning and information. Perhaps, for some purposes, these distinctions can be ignored. Perhaps, for some purposes, they should be ignored. You cannot, however, build a science of knowledge, a cognitive science, and ignore them. For knowledge is knowledge of the truth. That is why, no matter how fervently you might believe it, you cannot know that Paris is the capital of Italy, that pigs can fly or that there is a Santa Claus. You can, to be sure, put these "facts", these false sentences, into a computer's data base (or a person's head for that matter), but that doesn't make them true. It doesn't make them information. It just makes them sentences that, given the machine's limitations (or the person's ignorance), the machine (or person) treats as information.   But you can't make something true by thinking it is true, and you can't make something into information by regarding it as information.

So something—e.g., the sentence "Pigs can fly"--can mean pigs can fly without carrying that information. Indeed, given the fact that pigs can't fly, nothing can carry the information that pigs can fly. This is why, as commonly understood, information is such an important, such a useful, commodity. It gives you what you need to know—the truth. Meaning doesn't.

Information (once again, as it is commonly conceived) is something closely related to what natural signs and indicators provide. We say that the twenty rings in the tree stump indicate, they signify, that the tree is twenty years old. That is the information (about the age of the tree) the rings carry. We can come to know how old the tree is by counting the rings. Likewise, the rising mercury in a glass tube, a thermometer, indicates that the temperature is rising. That is what the increasing volume of the mercury is a sign of. That is the information the expanding mercury carries and, hence, what we can come to know by using this instrument. We sometimes use the word "meaning" to express this sentential content (what we can come to know) but this sense of the word, a sense of the word in which smoke means (indicates, is a sign of) fire, must be carefully distinguished from a linguistic sense of meaning in which the word "fire" (not the word "smoke" nor smoke itself) means fire. In a deservedly famous article, Paul Grice (1957) dubbed this informational kind of meaning, the kind of meaning in which smoke means (indicates, is a sign of) fire, <u>natural</u> <u>meaning</u>. With this kind of meaning, natural meaning, if an event, <u>e</u>, means (indicates, is a sign) that so-and-so exists, then so-and-so must exist. The red spots on her face can't mean, not in the natural sense of meaning, that she has the measles if she doesn't have the measles. If she doesn't have the measles, then perhaps all the spots mean in this natural sense is that she has been eating too many sweets. This contrasts with a language related (Grice called it "non-natural) meaning in which something (e.g., the sentence "She has the measles") can mean she has the measles even when she doesn't have them. If she doesn't have the measles, the sentence is false but that doesn't prevent it from meaning that she has the measles. If <u>e</u> (some event) means, in the natural sense,

that s̲ is F, however, then s̲ has to be F. Natural meaning is what indicators indicate. It is what natural signs are signs of. Natural meaning is information. It has to be true.

This isn't to say that we must know what things indicate, what information they carry. We may not know. We may have to find this out by patient investigation. But what we find out by patient investigation—that the tracks in the snow mean so-and-so or shadows on the film indicate such-and-such—is something that was true before we found it out. In this (natural) sense of meaning, we discover what things mean. We don't, as we do with linguistic or non-natural meaning, assign, create or invent it. By a collective change of mind we could change what the words "lightning" and "smoke" mean, but we cannot, by a similar change of mind, change what smoke and lightning mean (indicate). Maybe God (by changing natural laws) can, but we can't. What things mean, what they indicate, what information they provide, is in this way objective. It is independent of what we think or believe. It is independent of what we know. We may seek information in order to obtain knowledge, but the information we seek doesn't depend for its existence on anyone coming to know. It is, so to speak, out there in the world awaiting our use (or abuse) of it. Information is, in this way, different from knowledge. Information doesn't need conscious beings to exist, but knowledge does. Without life there is no knowledge (because there is nobody to know anything), but there is still information. There still exists that which, if knowers existed, they would need to know.

2. Information and Communication.

If this is, even roughly, the target we are aiming at, the idea of information we want a theory of, then a theory of information should provide some systematic, more

precise, perhaps more analytical, way of thinking about this epistemologically important commodity. If possible, we want a framework, a set of principles, that will illuminate the nature and structure of information and, at the same time, reveal the source of its power to confer knowledge on those who possess it.

In Dretske (1981, 1983) I found it useful to use Claude Shannon's Mathematical Theory of Communication (1948) for these purposes (see also Cherry 1957 for a useful overview and Sayre 1965 for an early effort in this direction).   Shannon's theory does not deal with the semantic aspects of information. It has nothing to say about the news, message, or content of a signal, the information (that the enemy is coming by sea, for instance) expressed in propositional form that a condition (a lantern in a tower) conveys. It does, however, focus on what is, for epistemological purposes, the absolutely critical relation between a source of information (the enemy's position) and a signal (a lantern in the tower) that carries information about that source. Shannon's theory doesn't concern itself with what news, message or information is communicated from s (source) to r (receiver) or, indeed, whether anything intelligible is communicated at all. As far as Shannon's theory is concerned, it could all be gibberish (e.g., "By are they sea coming."). What the theory does focus on in its theory of mutual information (a measure of amount of information at the receiver about a source) is the question of the amount of statistical dependency existing between events occurring at these two places. Do events occurring at the receiver alter in any way the probability of what occurred at the source? Given the totality of things that occur, or that might occur, at these two places, is there, given what happens at the receiver, a reduction in (what is suggestively called) the uncertainty of what happened at the source?

This topic, the communication channel between source and receiver, is a critically important topic for epistemology because "receiver" and "source" are just information-theoretic labels for knower and known. Unless a knower (at a receiver) is connected to the facts (at a source) in an appropriate way, unless there is a suitably reliable channel of communication between them, the facts cannot be known. With the possible exception of the mind's awareness of itself (introspection)--there is always, even in proprioception, a channel between knower and known, a set of conditions on which the communication of information—and therefore the possibility of knowledge—depends.   What we can hope to learn from communication theory is what this channel must look like, what conditions must actually exist, for the transmission of the information, needed to know

At one level, all this sounds perfectly familiar and commonplace. If someone cuts the phone lines between you and me, we can no longer communicate. I can no longer get from you the information I need in order to know when you are planning to arrive. Even if the phone lines are repaired, a faulty connection can generate so much "noise" (another important concept in communication theory) that not enough information gets through to be of much use. I hear you, yes, but not well enough to understand you. If we don't find a better, a clearer, channel over which to communicate, I will never find out, never come to know, when you plan to arrive.

That, as I say, is a familiar, almost banal, example of the way the communication of information is deemed essential for knowledge. What we hope to obtain from a theory of communication, if we can get it, is a systematic and illuminating generalization of the intuitions at work in such examples. What we seek, in its most general possible form, whether the communication occurs by phone, gesture, speech, writing, smoke signals, or

mental telepathy, is what kind of communication channel must exist between you and me

for me to learn what your plans are.   Even more generally, for any A and B, what must

the channel, the connection, between A and B be like for someone at A to learn something

about B?

The Mathematical Theory of Communication doesn't answer this question, but it

does supply a set of ideas, and a mathematical formalism, from which an answer can be

constructed. The theory itself deals in amounts of information, how much (on average)

information is generated at source $\underline{s}$ and how much (on average) information there is at

receiver $\underline{r}$ about this source. It does not try to tell us <u>what</u> information is communicated

from $\underline{s}$ to $\underline{r}$ or even, if some information is communicated, how much is enough to know

what is happening at $\underline{s}$. It might tell us that there are 8 bits of information generated at $\underline{s}$

about, say, the location of a chess piece on a chessboard (the piece is on KB-3) and that

there are 7 bits of information at $\underline{r}$ about the location of this piece, but it does not tell us

what information this 7 bits is the measure of nor whether 7 bits of information is enough

to know where the chess piece is. About that it is silent.

3. Using Communication Theory

We can, however, piece together the answers to these questions out of the elements

and structure provided by communication theory. To understand the way this might work

consider the following toy example (adapted from Dretske 1981) and the way it is handled

by communication theory. There are eight employees and one of them must perform some

unpleasant task. Their employer has left the job of selecting the unfortunate individual up

to the group itself, asking only to be informed of the outcome once the decision is made.

The group devises some random procedure that it deems fair (drawing straws, flipping a coin), and Herman is selected. A memo is dispatched to the employer with the sentence, "Herman was chosen" written on it.

Communication theory identifies the amount of information associated with, or generated by, the occurrence of an event with the reduction in uncertainty, the elimination of possibilities, represented by that event. Initially there were eight eligible candidates for the task. These eight possibilities, all (let us assume) equally likely, were then reduced to one by the selection of Herman. In a certain intuitive sense of "uncertainty", there is no longer any uncertainty about who will do the job. The choice has been made. When an ensemble of possibilities is reduced in this way (by the occurrence of one of them), the amount of information associated with the result is a function of how many possibilities there were (8 in this case) and their respective probabilities (.125 for each in this case). If all are equally likely, then the amount of information (measured in <u>bits</u>) generated by the occurrence of one of these <u>n</u> possibilities, I_g, is the logarithm to the base 2 of <u>n</u> (the power to which 2 must be raised to equal <u>n</u>):

(1) $I_g = \log_2 \underline{n}$

Since we started with eight possibilities all of which were all equally likely, I_g is $\log_2 8 = 3$ bits. Had there been 16 instead of 8 employees, Herman's selection would have generated 4 bits of information—more information since there is a reduction of more uncertainty. [1]

---

[1] If the probabilities of selection are not equal (e.g., probability of Herman = 1/6, probability of Barbara = 1/12, etc.), then I_g (average amount of information generated by the selection of an employee) is a weighted average of the information generated by the selection of each. I pass over these complications here since they aren't relevant to the use of communication theory in epistemology. What is relevant to epistemology is not how much information is generated by the occurrence of an event, or how much (on average) is generated by the occurrence of an ensemble of events, but how much of that information is transmitted to a potential knower at some receiver.

The quantity of interest to epistemology, though, is not the information <u>generated</u> by an event, but the amount of information <u>transmitted</u> to some potential knower, in this case the employer, about the occurrence of that event. It doesn't make much difference how much information an event generates: 1 bit or 100 gigabytes. The epistemologically important question is: how much of this information is transmitted to, and subsequently ends up in the head of, a person at <u>r</u> seeking to know what happened at <u>s</u>. Think, therefore, about the note with the name "Herman" on it lying on the employer's desk. How much information does this piece of paper carry about what occurred in the other room? Does it carry the information that Herman was selected? Would the employer, upon reading (and understanding) the message, know who was selected? The sentence written on the memo does, of course, mean in that non-natural or linguistic sense described above that Herman was selected. It certainly would cause the employer to <u>believe</u> that Herman was selected. But these aren't the questions being asked. What is being asked is whether the message indicates, whether it means in the natural sense, whether it carries the information, that Herman was selected. Would it enable the employer to <u>know</u> that Herman was selected? Not every sentence written on a piece of paper carries information corresponding to its (non-natural) meaning. "Pigs can fly" as it appears on this (or, indeed, any other) page doesn't carry the information that pigs can fly. Does the sentence "Herman was selected" on the employees' memo carry the information that Herman was selected? If so, why?

Our example involves the use of an information-carrying signal—the memo to the employer—that has linguistic (non-natural) meaning, but this is quite irrelevant to the way the situation is analyzed in communication theory. To understand why, think about an analogous situation in which non-natural (linguistic) meaning is absent. There are eight

mischievous boys and a missing cookie. Who took it? Inspection reveals cookie crumbs on Junior's lips. How much information about the identity of the thief do the crumbs on Junior's lips carry? For informational purposes, this question is exactly the same as our question about how much information about which employee was selected the memo to the employer carries. In the case of Junior, the crumbs on his lips do not have linguistic meaning. They have a natural meaning, yes. They mean (indicate) he took the cookie. But they don't have the kind of conventional meaning associated with a sentence like, "Junior took the cookie."

Communication theory has a formula for computing amounts of transmitted (it is sometimes called <u>mutual</u>) information. Once again, the theory is concerned not with the conditional probabilities that exist between particular events at the source (Herman being selected) and the receiver (Herman's name appearing on the memo) but with the average amount of information, a measure of the <u>general</u> reliability of the communication channel connecting source and receiver. There are eight different conditions that might exist at <u>s</u>: Barbara is selected, Herman is selected, etc. There are eight different results at <u>r</u>: a memo with the name "Herman" on it, a memo with the name "Barbara" on it, and so on. There are, then, sixty four conditional probabilities between these events: the probability that Herman was selected given that his name appears on the memo:

Pr[Herman was selected/the name "Herman" appears on the memo];

the probability that Barbara was selected given that the name "Herman" appears on the memo:

Pr[Barbara was selected/the name "Herman" appears on the memo];

and so on for each of the eight employees and each of the eight possible memos. The

transmitted information, I_t, is identified with a certain function of these 64 conditional

probabilities. One way to express this function is to say that the amount of information

transmitted, I_t, is the amount of information generated at s, I_g, minus a quantity called

equivocation, E, a measure of the statistical independence between events occurring at s

and r. [2]

(2) $I_t = I_g - E$

The mathematical details are not really important. A few examples will illustrate

the main ideas. Suppose the employees and messenger are completely scrupulous. Memos

always indicate exactly who was selected, and memos always arrive on the employer's

desk exactly as they were sent. Given this kind of reliability, the conditional probabilities

are all either 0 or 1.

Pr[Herman was selected/the name "Herman" appears on the memo] = 1

Pr[Barbara was selected/the name "Herman" appears on the memo] = 0

Pr[Nancy was selected/the name "Herman" appears on the memo] = 0

.

.

.

Pr[Barbara was selected/the name "Barbara" appears on the memo] = 1

Pr[Herman was selected/the name "Barbara" appears on the memo] = 0

Pr[Nancy was selected/the name "Barbara" appears on the memo] = 0

---

[2] Equivocation, E, is the weighted (according to its probability of occurrence) sum of individual contributions, $E(r_1)$, $E(r_2)$, . . . to equivocation of each of the possible events (eight possible memos) at r: $E = pr(r_1)E(r_1) + pr(r_2)E(r_2) + . . . pr(r_8)E(r_8)$ where $E(r_i) = -\sum pr(s_i/r_i) \bullet \log_2[pr(s_i/r_i)]$. If events at s and r are statistically independent then E is at a maximum $(E = I_g)$ and I_t is zero.

.

.

.

And so on for all employees and possible memos. Given this reliable connection, this trustworthy communication channel, between what happens among the employees and what appears on the memo to their employer, the <u>equivocation</u>, E turns out to be zero.[3] $I_t = I_g$: the memo on which is written an employee's name carries 3 bits of information about who was selected. All of the information generated by an employee's selection, 3 bits, reaches its destination.

Suppose, on the other hand, we have a faulty, a broken, channel of communication. On his way to the employer's office the messenger loses the memo. He knows it contained the name of one of the employees, but he doesn't remember which one. Too lazy to return for a new message, he selects a name of one of the employees at random, scribbles it on a sheet of paper, and delivers it. The name he selects happens, by chance, to be "Herman." Things turn out as before. Herman is assigned the job, and no one (but the messenger) is the wiser. In this case, though, the set of conditional probabilities defining equivocation (and, thus, amount of transmitted or mutual information) are quite different. Given that the messenger plucked a name at random, the probabilities look like this:

Pr[Herman was selected/the name "Herman" appears on the memo] = 1/8

Pr[Barbara was selected/the name "Herman" appears on the memo] = 1/8

.

.

Pr[Herman was selected/the name "Barbara" appears on the memo] = 1/8

Pr[Barbara was selected/the name "Barbara" appears on the memo] = 1/8

.

.

.

The statistical function defining equivocation (see footnote 2) now yields a maximum value of 3 bits. The amount of transmitted information, formula (2), is therefore zero.

These two examples represent the extreme cases: maximum communication and zero communication. One final example of an intermediate case and we will be ready to explore the possibility of applying these results in an information-theoretic account of knowledge.

Imagine the employees solicitous about Barbara's delicate health. They agree to name Herman on their note if, by chance, Barbara should be the nominee according to their random selection process. In this case I_g, the amount of information generated by Herman's selection would still be 3 bits: 8 possibilities, all equally likely, reduced to 1. Given their intention to protect Barbara, though, the probabilities defining transmitted information change. In particular

Pr[Herman was selected/the name "Herman" appears on the memo] = 1/2

Pr[Barbara was selected/the name "Herman" appears on the memo] = 1/2

The remaining conditional probabilities stay the same. This small change means that E, the average equivocation on the channel, is no longer 0. It rises to .25. Hence, according

---

[3] Either pr(s/r) = 0 or $\log_2[\text{pr(s/r)}] = 0$ in the individual contributions to equivocation (see footnote 2). Note: $\log_2 1 = 0$.

to (2), I_t drops from 3 to 2.75. Some information is transmitted, but not as much as in the first case. Not as much information is transmitted as is generated by the selection of an employee (3 bits)

This result seems to be in perfect accord with ordinary intuitions about what it takes to know. For it seems right to say that, in these circumstances, anyone reading the memo naming Herman as the one selected could not learn, could not come to know, on the basis of the memo alone, that Herman actually was selected. Given the circumstances, the person selected might have been Barbara. So it would seem that communication theory gives us the right answer about when someone could know. One could know that it was Herman in the first case, when the message contained 3 bits of information—exactly the amount generated by Herman's selection —and one couldn't know in the second and third case, when the memo contains 0 bits and 2.75 bits of information, something less than the amount generated by Herman's selection. So if information is what it takes to know, then it seem correct to conclude that in the first case the information that Herman was selected was transmitted and in the second and third case it was not. By focusing on the <u>amount</u> of information carried by a signal, communication theory manages to tell us something about the informational <u>content</u> of the signal—something about the news or message the signal actually carries—and, hence, something about what (in propositional form) can be known.

4. The Communication Channel

Let us, however, ask a slightly different question. We keep conditions the same as in the third example (Herman will be named on the memo if Barbara is selected), but ask whether communication theory gives the right result if someone else is selected. Suppose

Nancy is selected, and a memo sent bearing her name. Since the general reliability of the

communication channel remains exactly the same, the amount of transmitted information

(a quantity that, by averaging over all possible messages, is intended to reflect this general

reliability) also stays the same: 2.75 bits. This is, as it were, a 2.75 bit channel, and this

measure doesn't change no matter which particular message we happen to send over this

channel. If we use this as a measure of how much information is carried by a memo with

Nancy's name on it, though, we seem to get the wrong result. The message doesn't carry

as much information, 3 bits, as Nancy's selection generates. So the message doesn't carry

the information that Nancy was selected. Yet, a message bearing the name "Nancy" (or,

indeed, a memo bearing the name of any employee except "Herman") is a perfectly

reliable sign of who was selected. The name "Nancy" indicates, it means (in the natural

sense) that Nancy was selected even though a memo bearing the name "Herman" doesn't

mean that Herman was selected. The same is true of the other employees. The only time

the memo is equivocal (in the ordinary sense of "equivocal") is when it bears the name

"Herman." Then it can't be trusted. Then the nominee could be either Herman or

Barbara. But as long as the message doesn't carry the name "Herman" it is an absolutely

reliable indicator of who was selected. So when it bears the name "Nancy" ("Tom" etc.)

why doesn't the memo, contrary to communication theory, carry the information that

Nancy (Tom, etc.) was selected? A 2.75 bit channel is a reliable enough channel—at least

sometimes, when the message bears the name "Nancy" or "Tom," for instance--to carry a

3 bit message.

Philosophical opinions diverge at this point. Some are inclined to say that

Communication Theory's concentration on averages disqualifies it for rendering a useful

analysis of when a signal carries information in the ordinary sense of information. For,

according to this view, a message to the employer bearing the name "Nancy" does carry

information about who was selected. It enables the employer to know who was selected

even though he might have been misled had a message arrived bearing a different name.

The fact that the average amount of transmitted information (2.75 bits) is less than the

average amount of generated information (3 bits) doesn't mean that a particular signal

(e.g., a memo with the name "Nancy" on it) can't carry all the information needed to know

that Nancy was selected. As long as the signal indicates, as long as it means in the natural

sense, that Nancy was selected, it is a secure enough connection (channel) to the facts to

know that Nancy was selected even if other signals (a memo with the name "Herman" on

it) fail to be equally informative. Communication Theory, in so far as it concentrates on

averages, then, is irrelevant to the ordinary, the epistemologically important, sense of

information. [4]

Others will disagree. Disagreement arises as a result of different judgments about

what it takes to know and, therefore, about which events can be said to carry information

in the ordinary sense of information. The thought is something like this: a communication

channel that is <u>sometimes</u> unreliable is not good enough to know even when it happens to

be right. A channel of the sort described here, a channel that (unknown to the receiver)

sometimes transmits misleading messages, is a channel that should never be trusted. If it

is trusted, the resulting belief, even it happens to be true, does not possess the "certainty"

characteristic of knowledge. If messages are trusted, if the receiver actually believes that

---

[4] This is the view I took in Dretske (1981) and why I argued that the statistical functions of epistemological
importance were not those defining average amounts of information (equivocation, etc.), but the amount of
information associated with particular signals. It was not, I argued, average equivocation that we needed to
be concerned with, but the equivocation associated with particular signals (see Dretske 1981, 25-26).

Nancy was selected on the basis of a message bearing the name "Nancy," the resulting

belief does not, therefore, add up to knowledge. To think otherwise is like supposing that

one could come to know by taking the word of a chronic liar just because he happened, on

this particular occasion, and quite unintentionally, to be speaking the truth.

Imagine a Q meter designed to measure values of Q. Unknown to its users, it is

perfectly reliable for values below 100, but unpredictably erratic for values above 100. Is

such an instrument one that a person, ignorant of the instrument's eccentric disposition[5],

could use to learn values of Q below 100? Would a person who took a reading of "84" at

face value, a person who was caused to believe that Q was 84 by a reading of "84" on this

instrument, <u>know</u> that Q was 84? Does the instrument deliver information about values of

Q below 100 to trusting users? If your answer to these questions is "No," you are using

something like communication theory to guide your judgments about what is needed to

know and, hence, about when information is communicated. This instrument doesn't

deliver what it takes to know (i.e., information in the ordinary sense) because although the

particular reading ("84") one ends up trusting is within the instrument's reliable range (the

instrument wouldn't read "84" unless Q was 84) you don't know this. You would have

trusted it even if it had registered "104". The method being used to "track" the truth (the

value of Q) doesn't track the truth throughout the range in which that method is being

used.[6]

---

[5] If users were aware of the instrument's limited reliability, of course, they could compensate by ignoring readings above 100 and, in effect, make the instrument completely accurate in the ranges it was used (i.e., trusted). Practically speaking, this represent a change in the communication channel since certain readings (those above 100) would no longer be regarded as information-bearing signals.

[6] This way of putting the point is meant to recall Robert Nozick's (1981) discussion of similar issues. If the method being used to "track" (Nozick's term) the truth is insensitive to ranges of unreliability, then the method is not such as

Externalism is the name for an epistemological view that maintains that some of the conditions required to know that P may be, and often are, completely beyond the ken of the knower. You can, in normal illumination, see (hence, know) what color the walls are even if you don't know (because you haven't checked) that the illumination is normal. Contrary to Descartes, in normal circumstances you can know you are sitting in front of the fireplace even if you don't know (and can't show) the circumstances are normal, even if you don't know (and can't show) you are not dreaming or being deceived by some deceptive demon. According to externalism, what is important for knowledge is not that you <u>know</u> perceptual conditions are normal (the way they are when things are the way they appear to be), but that conditions actually <u>be</u> normal. If they are, if illumination (perspective, eyesight, etc.) are as you (in ordinary life) routinely take them to be, then you can see—and, hence, know—that the walls are blue, that you are sitting in front of the fireplace, and that you have two hands. You can know these things even if, for skeptical reasons, you cannot verify (without arguing in a circle) that circumstances are propitious. Information-theoretic accounts of knowledge are typically advanced as forms of externalism. The idea is that the information required to know can be obtained from a signal without having to know that the signal from which you obtain this information actually carries it. What matters in finding out that Nancy was selected (or in coming to know any other empirical matter of fact) is not that equivocation on the channel (connecting knower and known) be known to be zero. What is crucial is that it actually—whether known or not—<u>be</u> zero. This dispute about whether a memo bearing the name "Nancy" carries the information that Nancy was selected is really a dispute among

---

to satisfy the counterfactual conditions Nozick uses to define tracking. One would (using that method) have believed P even when P was false. See, also, Goldman's (1976) insightful discussion of the importance of distinguishing the ways we come to know.

externalists not about what has to be known about a communication channel for it to carry information. Externalists will typically agree that nothing has to be known. It is, instead, a dispute about exactly what (independently of whether or not it is known) constitutes the communication channel. In calculating equivocation between source and receiver—and, therefore, the amount of information a signal at the receiver carries about a source, should we count every signal that would produce the same resulting belief—the belief (to use our example again) that Nancy was selected? In this case we don't count memos carrying the name "Herman" since although these memos will produce false belief, they will not produce a false belief about <u>Nancy's</u> selection. If we do this, we get an equivocation-free channel. Information transmission is optimal. Or should we count every signal that would produce a belief about who was selected—whether or not it is Nancy? Then we count memos carrying the name Herman, and the communication channel, as so defined, starts to get noisy. The amount of mutual information, a measure of the amount of information transmitted, about who was selected is no longer equal to the amount of information generated. Memos—even when they carry the name "Nancy"--do not carry as much information as is generated the choice of Nancy because equivocal messages bearing the name "Herman" are used to reckon the channel's reliability even when it carries the message "Nancy." Or—a third possible option--in reckoning the equivocation on a communication channel, should we (as skeptics would urge) count <u>any</u> belief that would be produced by any memo (or, worse, any signal) whatsoever? If we start reckoning equivocation on communication channels in that way, then, given the mere possibility of misperception, no communication channel is <u>ever</u> entirely free of equivocation. The required information is <u>never</u> communicated. Nothing is known.

I do not—not <u>here</u> at least—take sides in this dispute. I merely describe a choice point for those interested in pursuing an information-theoretic epistemology. The choice one makes here—a choice about what collection of events and conditions are to determine the channel of communication between knower and known--is an important one. In the end, it determines what conclusions one will reach about such traditional epistemological problems as skepticism and the limits of human knowledge. I refer to this as a "choice" point to register my own belief that communication theory, and the concept of information it yields, does not solve philosophical problems. It is, at best, a tool one can use to express solutions—choices--reached by other means.

5. Residual Problems and Choices

What follows are three more problems or, as I prefer to put it, three more choices confronting anyone developing an information-theoretic epistemology that is based, even if only roughly, on an interpretation of information supplied by communication theory. I have my own ideas about which choices should be made and I will so indicate, but I will not here argue for these choices. That would require a depth of epistemological argument that goes beyond the scope of this paper.

A. <u>Probability</u>: In speaking of mutual information within the framework of communication theory, we imply that there is a set of conditional probabilities relating events at source and receiver. If these conditional probabilities are objective, then the resulting idea of information is objective. If they are subjective, somehow dependent on what we happen to believe, on our willingness to bet, on our level of confidence, then the resulting notion of information is subjective. If information is objective, then to the extent

that knowledge depends on information, knowledge will also be objective. Whether a

person who believes that P knows that P will depend on how, objectively speaking, that

person is connected to the world. It will depend on whether the person's belief (assuming

it is true) has appropriate informational credentials—whether, that is, it (or the evidence

on which it is based) stands in suitable probabilistic relations to events at the source. That

will be an objective matter, a matter to be decided by objective facts defining information.

It will not depend on the person's (or anyone else's) opinion about these facts, their level

of confidence, or their willingness to bet. If, on the other hand, probability is a reflection

of subjective attitudes, if the probability of e (some event at a source) given r (an event at

a receiver) depends on the judgments of people assigning the probability, then knowledge,

in so far as it depends on information, will depend on these judgments. Whether S knows

that P will depend on who is saying S knows that P.

I have said nothing here about the concept of probability that figures so centrally in

communication theory. I have said nothing because, as far as I can see, an information-

theoretic epistemology is compatible with different interpretations of probability.[7] One

can interpret it as degree of rational expectation (subjective), or (objectively) as limiting

frequency or propensity. In developing my own information-based account of knowledge

in (Dretske 1981) I assumed (without arguing for) an objective interpretation. There are, I

think, strong reasons for preferring this approach, but strictly speaking, this is optional.

The probabilities can be given a subjective interpretation with little or no change in the

formal machinery. What changes (for the worse, I would argue) are the epistemological

consequences.

---

[7] But see Loewer (1983) for arguments that there is no extant theory of probability that will do the job.

If probability is understood objectively, an informational account of knowledge takes on some of the characteristics of a causal theory of knowledge.[8] According to a causal theory of knowledge, a belief qualifies as knowledge only if the belief stands in an appropriate causal relation to the facts. I know Judy left the party early, for instance, only if her early departure causes me to believe it (either by my seeing her leave or by someone else—who saw her leave early—telling me she left). Whether my belief that she left early is caused in the right way is presumably an objective matter. It doesn't depend on whether I or anyone else know it was caused in the right way. For this reason everyone (including me) may be wrong in thinking that I (who believes Judy left early) know she left early. Or everyone (including me) may be wrong in thinking I don't know she left early. Whether or not I know depends on facts, possibly unknown, about the causal etiology of my belief. If probability is (like causality) an objective relation between events, then an information-theoretic account of knowledge has the same result. Whether or not someone knows is a matter about which everyone (including the knower) may be ignorant. To know whether S knows something —that Judy left early, say—requires knowing whether S's belief that Judy left early meets appropriate informational (i.e., probabilistic) conditions, and this is a piece of knowledge that people (including S herself) may well not have.

If, on the other hand, probability is given a subjective interpretation, information —and therefore the knowledge that depends on it—takes on a more relativistic character. Whether or not S knows now depends on who is attributing the knowledge. It will depend on (and thus vary with) the attributor of knowledge because, presumably, the person who is attributing the knowledge will be doing the interpreting on which the probabilities and,

---

[8] Goldman (1967) gives a classic statement of this theory.

therefore, the information and, therefore, the knowledge depends. As a result, it will turn

out that you and I can both speak truly when you assert and I deny that S knows Judy left

early. Contextualism (see Cohn 1986, 1988, 1999; DeRose 1995; Feldman 1999; Heller

1999; Lewis 1996) in the theory of knowledge is a view that embraces this result.

       B. <u>Necessary Truths:</u> Communication theory defines the amount of transmitted

information between source and receiver in terms of the conditional probabilities between

events that occur, or might have occurred, at these two places. As long as what occurs at

the source generates information--as long, that is, as the condition existing at a source is a

<u>contingent</u> state of affairs (a state of affairs for which there are possible alternatives) there

will always be a set of events (the totality of events that might have occurred there) over

which these probabilities are defined. But if the targeted condition is one for which there

are no possible alternatives, a <u>necessary</u> state of affairs, no information is generated. Since

a necessary state of affairs generates zero information, every other state (no matter how

informationally impoverished it might be) carries an amount of information (i.e., ≥ 0 bits)

needed to know about its existence. According to communication theory, then, it would

seem that nothing (in the way of information) is needed to know that 3 is the cube root of

27. Or, to put the same point differently, informationally speaking anything whatsoever is

good enough to know a necessary truth. Bubba's assurances are good enough to know that

3 is the cube root of 27 because his assurances carry all the information generated by that

fact. Mathematical knowledge appears to be cheap indeed.

       One way to deal with this problem is to accept a subjective account of probability.

The village idiot's assurances that 3 is the cube root of 27 need not carry the information

that 3 is the cube root of 27 if probability is a measure of, say, one's willingness to bet or one's level of confidence. On this interpretation, the probability that 3 is the cube root of 27, given (only) Bubba's assurances, may be anything between 0 and 1. Whether or not I know, on the basis of Bubba's assurances, that 3 is the cube root of 27, will then depend on how willing I am to trust Bubba. That will determine whether Bubba is a suitable informant about mathematics, a suitable channel for getting information about the cube root of 27.

Another way to deal with this problem is to retain an objective interpretation of probability but insist that the equivocation on the channel connecting you to the facts, the channel involving (in this case) Bubba's pronouncements, is to be computed by the entire set of things Bubba might say (on all manner of topics), not just what he happened to say about the cube root of 27.   If equivocation (and, thus, amount of transmitted information) is computed in this way, then whether or not one receives information about the cube root of 27 from Bubba depends on how generally reliable Bubba is. Generally speaking, on all kinds of topics, is Bubba a reliable informant? If not, then whether or not he is telling the truth about the cube root of 27, whether or not he could be wrong about that, he is not a purveyor of information. One cannot learn, cannot come to know, that 3 is the cube root of 27 from him. If Bubba is a generally reliable informant, on the other hand, then he is someone from whom one can learn mathematics as well as any other subject about which he is generally reliable.

A third way to deal with the problem, the way I took in Dretske 1981, is to restrict one's theory of knowledge to perceptual knowledge or (more generally) to knowledge of contingent (empirical) fact. Since a contingent fact is a fact for which there are possible

alternatives, a fact that might not have been a fact, a fact that (because it has a probability less than one) generates information, one will always have a channel of communication between knower and known that is possibly equivocal, a channel that might mislead. If a theory of knowledge is a theory about this limited domain of facts, a theory (merely) of <u>empirical</u> knowledge, then communication theory is prepared to say something about an essential ingredient in such knowledge. It tells you what the channel between source and receiver must be like for someone at the receiver to learn, come to know, empirical facts about the source.

     C. <u>How Much Information is Enough?</u> I have been assuming that information is necessary for knowledge. The employer can't know who was selected—that it was Herman--unless he receives the required information. Following a natural line of thought, I have also been assuming that if information is understood in a communication-theoretic sense, then the amount of information received about who was selected has to be equal to (or greater) than the amount of information generated by the selection. So if Herman's selection generates 3 bits of information (there are eight employees, each of which has an equal chance of being selected), then to know who was selected you have to receive some communication (e.g., a message with the name "Herman" on it) that carries at least that much information about who was selected. If it carries only 2.75 bits of information, as it did in the hypothetical case where employees were determined to protect (i.e., not name) Barbara, then the message, although enough (if it carries the name "Herman") to produce true belief, could not produce knowledge. In order to know what happened at <u>s</u> you have

to receive as much information—in this case 3 bits—about s̲ as is generated by the event you believe to have occurred there.

My examples were deliberately chosen to support this judgment. But there are other examples, or other ways of framing the same example, that suggest otherwise. So, for instance, suppose the employees' messages are not so rigidly determined. Messages bearing the name "Herman" make it 99% probable that Herman was selected, messages bearing the name "Barbara" make it 98% probable that Barbara was chosen, and so on (with correspondingly high probabilities) for the remaining six employees. As long as these probabilities are neither 0 nor 1, the individual contributions to equivocation (see footnote 2) will be positive. The equivocation, E, on the channel will, therefore, be greater than 0 and the amount of transmitted information will be less than the amount of information generated. Messages about an employee's selection will never carry as much information as is generated by that employee's selection. Full and complete information about who was selected, the kind of information (I have been arguing) required to know who was selected, will never be transmitted by these messages. Can this be right? Is it clear that messages sent on this channel do not carry the requisite information? Why can't the employer know Herman was selected if he receives a memo with the name "Herman" on it? The probability is, after all, .99.

If a probability of .99 is not high enough, we can make the equivocation even less and the amount of information transmitted even greater by increasing probabilities. We can make the probability that X was selected, given that his or her name appears on the memo, .999 or .9999. As long as this probability is less than 1, equivocation is positive and the amount of transmitted information less than information generated. Should we

conclude, though, that however high the probabilities become, as long as $E > 0$ and, therefore, $I_t < I_g$), not enough information is transmitted to yield knowledge? If we say this, doesn't this make the informational price of knowledge unacceptably high? Isn't this an open embrace of skepticism?

If, on the other hand, we relax standards and say that enough information about conditions at a source is communicated to know that what condition exists there even when there is a permissibly small amount of equivocation, what is permissibly small? If, in order to know that Herman was selected, we don't need all the information generated by his selection, how much information is enough?

Non-skeptics are tugged in two directions here. In order to avoid skepticism, they want conditions for knowledge that can, at least in clear cases of knowledge, be satisfied. On the other hand, they do not want conditions that are too easily satisfied else clear and widely shared intuitions about what it takes to know are violated. Reasonable beliefs, beliefs that are very probably true, are clearly not good enough. Most people would say, for instance, that if S is drawing balls at random from a collection of balls (100, say) only one of which is white, all the rest being black, you can't, before you see the color of the ball, know that S selected a black ball even though you know the probability of its being black is 99%. S might, for all you know, have picked the white ball. Things like that happen. Not often, but often enough to discredit a claim that (before you peek) you know it didn't happen on this occasion. Examples like this suggest that knowledge requires eliminating all (reasonable? relevant?) chances of being wrong, and elimination of these is simply another way of requiring that the amount of information received about the state known to exist be (at least) as much as the amount of information generated by that state.

There are different strategies for dealing with this problem. One can adopt a relativistic picture of knowledge attributions wherein the amount of information needed to know depends on contextual factors. In some contexts, reasonably high probabilities are enough. In other contexts, perhaps they are not enough. How high the probabilities must be, how much equivocation is tolerated, will depend on such things as how important it is to be right about what is occurring at the source (do lives depend on your being right or is it just a matter of getting a higher score on an inconsequential examination?), how salient the possibilities are of being wrong, and so on.

A second possible way of dealing with the problem, one that retains an absolute (i.e., non-relativistic) picture of knowledge, is to adopt a more flexible (I would say more realistic) way of thinking about the conditional probabilities defining equivocation and, therefore, amount of transmitted information. Probabilities, in so far as they are relevant to practical affairs, are always computed against a set of circumstances that are assumed to be fixed or stable. The conditional probability of s, an event at a source, given r, the condition at the receiver is really the probability of s, given r within a background of stable or fixed circumstances B. To say that these circumstances are fixed or stable is not to say that they cannot change. It is only to say that for purposes of reckoning conditional probabilities, such changes are set aside as irrelevant. They are ignored. If the batteries in a measuring instrument are brand new, then even if it is possible, even if there is a non-zero probability, that new batteries are defective, that possibility is ignored in calculating the amount of information the instrument is delivering about the quantity it is being used to measure. The non-zero probability that B fails, that the batteries are defective, does not contribute to the equivocation of instruments for which B holds, instruments whose batteries are functioning

well. The same is true of all communication channels. The fact—if it is a fact--that there

is a non-zero probability that there were hallucinatory drugs in my morning coffee, does

not make my current (perfectly veridical) experience of bananas in the local grocery store

equivocal. It doesn't prevent my perception of bananas from delivering the information

needed to know that they (what I see) are bananas. It doesn't because the equivocation of

the information delivery system, my perceptual system, is computed taking as given the de

facto condition (no hallucinatory drugs) of the channel. Possible (non-actual) conditions of

this channel are ignored even if there is a non-zero probability that they actually exist. The

communication of information depends on their being, in fact, a reliable channel between a

source and a receiver. It doesn't require that this reliability itself be necessary.

REFERENCES

Cherry, C. 1957. On Human Communication, Cambridge, MA: MIT Press

Cohen, S. 1986. Knowledge and Context. Journal of Philosophy 83:10

Cohen, S. 1988. How to Be a Fallibilist. Philosophical Perspectives, Vol. 2, J. Tomberlin, ed.
    Atascadero, CA; Ridgeview Publishing

Cohen, S. 1991. Scepticism, Relevance, and Relativity. In Dretske and His Critics. Cambridge,
    MA; Blackwell: 17-37.

Cohen, S. 1999. Contextualism, Skepticism, and the Structure of Reasons. In Tomberlin (1999):
    pp. 57-90

DeRose, K. 1995. Solving the Skeptical Problem. Philosophical Review, 104.1: 1-52

Dretske, F. 1981. Knowledge and the Flow of Information, Cambridge, MA: MIT Press.

Dretske, F. 1983. 'Multiple book review of Knowledge and the Flow of Information', Behavioral

    and Brain Sciences 6 (1): 55-89.

Feldman, R. 1999. Contextualism and Skepticism. In Tomberlin (1999): 91-114.

Goldman, A. 1967. A Causal Theory of Knowing. Journal of Philosophy, 64: 357-72.

Goldman, A. 1976. Discrimination and Perceptual Knowledge. Journal of Philosophy, 73: 771-

    791.

Heller, M. 1999. The proper role for contextualism in an anti-luck epistemology. In Tomberlin

    (1999): 115-130.

Lewis, D. 1996. Elusive Knowledge. Australian Journal of Philosophy, 74.4: 549-567

Loewer, B. 1983. Information and Belief. Behavioral and Brain Sciences 6: 75-76,

Nozick, R. 1981. Philosophical Explanations. Cambridge, MA; Harvard University Press.

Sayre, K. 1965. Recognition: A Study in the Philosophy of Artificial Intelligence, South Bend,

    IN: University of Notre Dame Press.

Shannon, C. 1948. The mathematical theory of communication. <u>Bell Systems Technical Journal</u>

27: 379-423, 623-56; reprinted with introduction by W. Weaver, Urbana, IL: The University

of Illinois Press, 1949.