

Chapter Seven – Sampling & Analogical Reasoning

“We come to believe that the results of opinion polls
are what people believe, as if our beliefs
can be encapsulated in such sentences as
‘I approve’ and ‘I disapprove.’”
-- Neil Postman

“Analogies prove nothing, that is quite true,
but they can make one feel more at home.”
--Sigmund Freud

Two of the more important kinds of inductive reasoning are sampling and analogical reasoning. Sampling is a type of simple induction that generalizes from what is known about some individuals of the same type (a sample) to others of that type. Analogical reasoning involves comparing things on the basis of similarities.

1. Surveys and Polls

Surveys and polls have become a part of everyday life in our country. How are they done and why does anyone consider them reliable sources of useful information? Let’s take a close look at part of the National Annenberg Election Survey (NAES) published on May 17, 2004, by the Annenberg Public Policy Center of the University of Pennsylvania. The survey asked people for their opinion on same-sex marriages, which became legal in Massachusetts on the day the results of the poll were released. They also inquired into people’s opinion on a constitutional amendment that would prohibit such marriages. Their press release began with the following statements.

As same-sex marriage begins today in Massachusetts, Americans remain opposed to the concept but still dubious about prohibiting it by a constitutional amendment, the University of Pennsylvania’s National Annenberg Election Survey shows.

Interviewing of 3,775 adults from May 3 through 16 showed that 61 percent of the public said they opposed a law that would permit same-sex marriages in their state, while 30 percent said they favored it.

Those findings were not significantly different from the last Annenberg measure of the question. February 14 through 23 polling found that 64 percent were opposed to a same-sex marriage law in their state while 30 percent were in favor.

Nor was there significant change on the issue of a constitutional amendment to prohibit states from allowing same-sex marriages. In the latest polling, 42 percent supported such an amendment while 50 percent opposed one. In the February 14-23 period (which ended one day before President Bush called for an amendment), 41 percent favored an amendment and 48 percent opposed one.

The NAES asserts that *Americans* are opposed to same-sex marriage and are dubious about the constitutional amendment. We’re told that 3,775 adults were interviewed. What justifies drawing a conclusion about *Americans*—who number about 300,000,000—from a **sample** of just 3,775?

First, we should note that *Americans* is short for “American adults,” which is considered the **target population** for this survey. Only American adults were in the sample, so the conclusion can only be about American adults. Even so, how could 3,775 adults be **representative** of the entire adult population? What kind of

method did the Annenberg folks use to give them the confidence to **generalize** about the adult population as a whole based on their sample of fewer than 4,000? We're told

The 2004 National Annenberg Election Survey is based on telephone interviews which began October 7, 2003 and will continue past Election Day.

The sample of telephone exchanges called was randomly selected by a computer from a complete list of thousands of active residential exchanges across the country. Within each exchange, random digits were added to form a complete telephone number, thus permitting access to both listed and unlisted numbers. Within each household, one adult was designated by a random procedure to be the respondent for the survey. The interviewing is conducted by Schulman, Ronca, Bucuvalas, Inc.

The results have been weighted to take account of household size and number of telephone lines into the residence and to adjust for variation in the sample relating to geographic region, sex, race, age and education.

This report deals with interviewing conducted from May 2 through May 16, when 3,775 people were asked about a proposed constitutional amendment. In theory, in 19 cases out of 20 the results will differ by no more than three percentage points, up or down, from what would have been obtained by interviewing all American adults. The question about a state law allowing same-sex marriages was asked of 1,181 people, and the margin of sampling error for those answers is plus or minus three percentage points. For smaller subgroups, the margin of sampling error would be higher.

In addition to sampling error, the practical difficulties of conducting any survey of public opinion may introduce other sources of error into the poll. Variations in the wording and order of questions, for example, may lead to somewhat different results.

Without getting bogged down in too many details regarding these procedures, we can see that the polling firm has a number of specific techniques it uses to get a fair and balanced cross section of the adult population in its sample. Yet, the pollsters admit that there are **sampling errors** and "practical difficulties" that mean, in effect, they can't guarantee their results. Nevertheless, they are quite confident that if someone were to repeat the interviews using their methods that 95% of the time (19 out of 20) they'd get the same results plus or minus three percentage points. *In theory*, they say, we can accept these results as what we'd have gotten had we interviewed all American adults.

We'll get to the notion of **sampling error** in due time. First, let's review the method of selecting the sample. The validity of sampling depends upon the **representativeness** of the sample, which in turn depends on *how* the sample was selected and *how many* were in the sample. A sample is likely to be representative if a method of selecting the sample is used that will result in **a good cross section of typical items in the target population**. If a **biased** method is used to select the sample, then the reasoning will not be justified. If the sample is *too small*, it is likely to be biased. We can never be absolutely sure a sample is representative, but the more unbiased the method of selecting the sample is, the more likely it will be representative. If a method is used that favors or disfavors the selection of particular items to the sample, then the sample will be biased. For example, a sample of American adults that favored the selection of Republicans over Democrats would be biased no matter how large it is.

A method of selecting a sample that gives each item in the target population an equal chance of being selected is considered the most unbiased way of sampling. Such a sample is called a **random sample**. Let's say that we manufacture widgets and want to make sure our widgets perform to our standards. We could test every widget that comes off the assembly line. Sometimes this is possible; sometimes it is necessary. This is the only

way we can be sure that every item performs exactly as it is supposed to perform. However, sometimes it is impossible to test every item, e.g., when the test would destroy or damage the item. Often, it is not necessary to test every item. Moreover, sometimes it is impractical to test every item, e.g., when the target population is extremely large. Testing an adequate random sample—rather than all items in a population—will be sufficient.

How do we select a random sample, and *how many* should we select? First, do not let the word ‘random’ mislead you. You might think that someone standing on the assembly line, whimsically selecting items, would be collecting a random sample. She wouldn’t. The mark of a **random sample** is **that each item in the target population has an equal chance of being selected**. You could not be sure that the person whimsically choosing items was in fact free of bias in her method of selection. She may be unaware of her bias and you may not be able to identify any bias. To be sure that the method of selection is unbiased, we should use another method. For example, we could program a machine to select items randomly by numbering each item and having the machine select random numbers.

If a random sample is impossible or impractical, we might use the method of **systematic sampling**. For example, we could have a robot select every 10th, 100th, or 1000th widget as it came off the assembly line. We could program the computer to select a random number (under 11, 101, or 1001) to determine which widget will be the first tested. A systematic sample would *not* be a truly random sample. It might even lead to a biased sample in some cases. Imagine, for example, an assembly line running 24 hours a day, 5 days a week. The company has three different work crews, each working an 8-hour shift. If the robot is programmed so that it continually skips one shift, or significantly under-represents one shift, the sample will be biased. We can get a good cross section of the widgets using the systematic method, but we must make sure the robot selects an adequate sample from each of the three shifts.

Some very famous surveys have been biased. The Kinsey report, for example, is the source for a homosexuality statistic widely reported in both the mass media and in scientific publications, though the statistic is based on biased samples and is erroneous. Kinsey’s famous studies on sexual behavior have been repeatedly cited as the basis for the claim that 10% of the population is gay. In fact, numerous studies have been done which put the percentage of *adults who describe themselves as exclusively gay* as between one and two percent.¹ The reason that the Kinsey statistic on gays was so much higher than other studies is due to the biased samples he used. (*Bias* refers to *unintentional inclination*, not to homophobia.) He gathered his data, in part, by distributing questionnaires to prisoners and to people who attended his lectures on sexuality, neither of which were likely to be a good cross section of Americans.

2. Inductive generalizations

Surveys and polls use inductive reasoning to draw conclusions about target populations based on samples taken from those populations. Justified inductive generalizations can provide us with a useful means of facing the future because they are *predictive*. They tell us not only about the people we have interviewed or the items we have observed and measured; they tell us about people we haven’t interviewed and about items we have not observed or measured.

The testing of causal hypotheses in science is based, in part, on the same inductive principle as polling: We need only study a *part* of a population to gain knowledge about that population in general. A scientist does not have to study *every* virus of a certain type in order to draw justifiable conclusions about that type of virus. Since the number of individual viruses of any given type is probably extremely large, it is possible that in some cases the percentage of individuals of a type actually observed might be close to zero. The same would be true for an astronomer studying any particular type of star. No matter how many stars astronomers actually observe, that number will be an infinitesimally small fraction of all stars. The same would be true for a chemist studying human DNA. The same is true for a pollster studying the opinions or behavior patterns (beliefs and actions) of human populations. For example, there are about 300,000,000 Americans. The poll which concluded that 61% of all American adults are opposed to same-sex marriages was based on a sample of 3,775 adults, which is a very small percentage of the total population. Each person polled represents about 80,000 others. However, this sample is immensely larger than any percentage of the “population” a chemist, biologist, or physicist is likely to study in a

lifetime of scientific investigation of molecules, cells, or atomic particles. The samples studied by the chemist, the biologist, and the physicist are generally homogeneous, however. Once the type of item to be studied is identified, selecting which items to study is not a major problem since “one water molecule is pretty much like any other water molecule,” etc. What matters is not what *percentage* of the total target population is in the sample. **What matters is how likely it is that the sample is a good cross section of the target population.** What matters is how likely it is that the items in the sample are *typical* of the target population.

More problematic than the size of polling samples is that opinions are treated as if they were observable, measurable, and fixed qualities. The “technique of polling promotes the assumption that an opinion is a thing inside people that can be exactly located and extracted by the pollster’s questions” (Postman 1992: 134). Furthermore, it promotes the assumption that we *ought* to have an opinion on the issue being polled. Maybe what we *ought* to have is information that would help us make a reasonable judgment about the issue. Maybe questions ought to stimulate thought and discussion about an issue rather than end it by gathering useless data from an uninformed populace and announcing some statistic.

3. The method of selecting the sample

To select its sample the Annenberg survey on same-sex marriages used a rather elaborate telephone survey method, involving randomly selected telephone exchanges (the first three digits of a local number) from across the country. Since the last four numbers of each person called were selected randomly, both listed and unlisted numbers were included. Finally, a random procedure was used to select which person in the household would be interviewed. As stated above, the ideal way of selecting a sample would be to use a method that gives every member of the target population an equal chance of being in the sample. This would give us a *random sample*. However, a true random sample is not possible or feasible in polling large populations. The method used in the Annenberg Poll, *telephone sampling*, is perhaps the most common method of conducting polls. Area codes and phone numbers can be randomly generated by a computer or randomly selected from lists of phone numbers, ensuring that a good cross section of the state or nation is called. It is true that such a method will *systematically exclude* large sections of the target population: people not at home, people who are home but who will not participate, and people without telephones. Ask yourself, however, whether such people are likely to be significantly different with respect to their beliefs and actions than people who are home when the pollsters call? How would you go about finding out such a thing? If people are not home, will not participate, or do not have a telephone how can you poll them to see if they are significantly different from the people who are home and do participate? The only thing we can be sure of is that people who do not have a telephone are never counted in telephone surveys.

Any other method than the telephone survey of a large target population would probably be too costly and time consuming. Door to door sampling is out of the question for a statewide survey by a polling firm. Randomly selecting addresses to poll in a town is possible and safe only in small towns. Using mailed questionnaires allows one to send out the survey to all members or a good cross section of a target population, but there is no guarantee that those who return the survey are representative. Such a method can be very costly, also. More importantly, *self-selected samples are always biased*.

Thus, at most, we should probably say that the Annenberg Poll uses a method that is likely to give us a good idea of what a good portion of society believes.

4. Size of sample and margin of error

Next, let’s consider the size of the Annenberg sample. How many were in the sample, and is it large enough to give us a good cross section of the target population?

You might think that the Poll was biased because it was based on the views of only 3,375 people out of a population of 300 million. Actually, most pollsters would consider this sample to be of adequate size. How a sample of one, two, or three thousand can be representative of a population of 30,000,000 or 300,000,000 is a curious thing and worthy of concern. I'll try to explain it as best I can, though the 'real' reason has something to do with Gaussian curves, standards of deviation, probability distributions, and a few other statistical concepts best left to experts to explain. I will begin my non-expert explanation with the concept of **margin of sampling error**.²

The Annenberg Poll states that it has a margin of sampling error of plus or minus three percentage points. This statistic is calculated by plugging the value of the size of the sample into a statistical formula. Many news articles refer to the margin of sampling error as the *margin of error*. As a result, this statistic may give the appearance of being a measure of the poll's accuracy, i.e., representativeness, but *margin of sampling error* is completely independent of representativeness. A small margin of error (plus or minus 3 percent or less) tells us that the researchers are extremely confident that if they were to repeat the telephone survey one hundred times, in 95 of those times they would get the same results *plus or minus the margin of sampling error*. In other words, the Annenberg pollsters are extremely confident that there is a high probability that the percentage of American adults who oppose same-sex marriage is between 58% and 64%.

If the poll were repeated, no one would expect each sample to result in exactly the same statistics. There would be some variation due to chance. The margin of sampling error gives us a good indication what that chance variation is likely to be. The statistical formula that measures margin of error measures only a statistical probability of chance differences among individual items in a random sample. It does not measure errors due to biased data, erroneous calculations, or inappropriate application of the statistical data. I repeat that it can therefore be a very misleading statistic. A small margin of error might be taken to indicate that a sample is very representative of its population. However, the margin of error is dependent solely on the *size* of the sample. If the *method* of selecting that sample is biased to begin with, a small margin of error won't help. In other words, a low margin of error might be said to mean that *if the sample is not representative, that fact is not likely due to the sample size being too small!*

In addition, the margin of error does not decrease by a one-to-one factor with sample size. Margin of error varies with the square root of the sample size. Thus, to increase the reliability of a properly done statistical study 10-fold, one would have to increase the size of the sample 100-fold (Langley 1970: 45).³ Often, increasing the accuracy of a properly done study by two or three percent would be extremely costly; hence, researchers commonly use samples of under 1,000. In fact, it is probably safe to say that the size of most public opinion polls is determined in large part by *cost*. If one can get a margin of error of about plus or minus 3% with a sample of about 1,200 why spend the money, time, and effort to double or triple the sample size, when the decrease in margin of error would be barely noticed or appreciated?

Another thing to consider is that when pollsters divide their sample into sub-groups, the sample sizes for each sub-group must be smaller than the sample of all those polled. Therefore, the margin of sampling error for sub-groups will be *greater* than the margin of error for all those polled. For example, in the Annenberg poll, the strongest opposition to same-sex marriage laws comes from those who attend church frequently and from the elderly. Among those who go to religious services more than once a week, 82 percent are opposed to same sex-marriage and 81 percent of those 65 and older oppose. The margin of sampling for the frequent church goers and the elderly is greater than plus or minus three percentage points.

Intuitively, each of us would recognize that if the Annenberg Poll surveyed only a half dozen Americans, the sample size would be too small to warrant justifying any beliefs about the target population as a whole. In addition, it might seem obvious that the greater the number of people interviewed, the greater the probability that a sample is representative of the entire population. A closer examination will reveal, however, that what seems obviously true is actually obviously false in most cases.

Remember what we said above about most scientific samples. No matter how large the sample gets when we are studying DNA, viruses, stars, etc., it will always equal a negligible percentage of all DNA, viruses, stars, etc. Many human populations are also very large, and pollsters will never be able to survey any more than a very small fraction of the target population. Increasing the size of the Annenberg sample to 10,000, for example, would increase its percentage of the total population by a negligible percentage. It is generally pointless to try to get a sample that would be the size of a large percentage of the target population. In fact, such large samples would defeat the purpose of sampling! We sample in order to be able to study only a small part of a population to gain

useful information about the whole population. If a target population is so small that we could realistically study nearly all the members of the population, then sampling would be unnecessary.

It might seem that there is a major difference between studying human populations for their opinions and studying DNA or stars for their properties. The objects of scientific study are homogeneous; whereas human beings are a rather heterogeneous lot when it comes to their opinions. Random differences among homogeneous items are statistically unimportant (Werkmeister 1948: 479). Pollsters study a very heterogeneous aggregate in studying the opinions and behaviors of human groups. Even so, if we select the sample properly, the large random differences expected in the opinions and behaviors of a human population should tend to cancel out one another. If the method of selecting the sample is proper, a relatively small sample can give reliable results. A poll with a sample as small as about 1,200 taken from a target population of 300,000,000 can have a margin of error of about plus or minus 3%. In fact, whether the target population has 3 million, 30 million or 300 million, the margin of error would be the same. The margin of error depends on the *size of the sample*, not on the size of the sample in proportion to the total target population. An example might help clarify this.

Imagine that you have a computer data bank with 30,000,000 items. The items are *numbers*. They are either *even* or *odd*, but are not sequential. Your task is to figure out how many of the items are even and how many are odd. If the numbers were sequential, you would not need to do any sampling. You would know that 50% are even and 50% are odd. However, since the numbers are not sequential, you can either examine 30,000,000 items or do a random sampling. Even if it only took one second to examine each item and record the results, it would take you about four years, working eight hours a day, five days a week, to complete your task. Your computer could examine each item in much less time, of course, but we need not overwork the computer, either. We could program it to select items randomly from the data bank and keep track of the oddness or evenness of each item.

After ten items had been selected, would you feel confident that the percentages of odds and evens selected is representative of the data bank? You shouldn't. Ten would be too few to base our judgment on. How about 100? Would that be enough? Well, it would clearly be better than ten. In fact, the margin of error would be reduced from about plus or minus 30% for a sample of 10, to about plus or minus 10% for a sample of 100. But a sample of 100 is still too small. How about a sample of 1,000? Wouldn't you feel confident that a computer randomly selecting items from the data bank would give you a good cross section of those items when it had selected 1,000 items? In fact, the margin of error is now down to about plus or minus 3%.

You might think that things would be a bit more complicated if the items in the data bank had more than two properties (*odd* or *even*). However, it would not make any difference to the representativeness of our sample. We could have had our computer identify not only whether the sample items were odd or even, but whether the sum of their digits was odd or even, and whether they were prime numbers, and whether they ended in 5, etc. A random sample of 1,000 examined for two properties will be as representative as a sample of 1,000 examined for more than two properties. So, whether we are sampling opinions with only two variables (e.g., *yes* or *no*) or several variables (e.g., *extremely concerned*, *somewhat concerned*, *not too much concerned*, *not at all concerned*), we can be equally confident of similarly sized samples selected by similar methods.

Obviously, a very small sample is likely to be unrepresentative, even if the population from which the sample is taken is very small. For very small populations, the safest method is to study all the members of the population. For example, if you had a box with three marbles in it and you picked out one black marble, even though the one marble represents one-third of the population, the probability that the next marble you select will be black is impossible to tell without more information. On the other hand, if you had 30,000,000 marbles in the box, you could make a very accurate prediction about the percentages of different colors after randomly selecting a very small percentage of the total. As noted above, with a sample of 1,000 the margin of error would be about plus or minus 3%. How much could you reduce the margin of error by doubling your sample? Maybe about plus or minus 1%. Would it be worth it? It might be very costly to increase the size of a poll to 2,000 from 1,000. Would the small reduction in margin of error be worth it?

4.1 The *Literary Digest* fiasco

A classic example of fallacious sampling occurred in 1936 when *Literary Digest* magazine concluded on the basis of a sample of 2,376,000 that Landon would defeat Franklin D. Roosevelt by 57% to 43%. Roosevelt won by a landslide, getting 62% of the votes. How did the pollsters err? Their sample was *biased*. The poll selected its sample from readers of the magazine, lists of registered automobile owners, and those with telephones. It systematically excluded the poor, a significant stratum of society, many of whom voted Democratic.

The *Literary Digest* poll could have been improved only by improving its *method of sample selection*.

“A question, even of the simplest kind, is not and can never be unbiased.” --Neal Postman

Increasing the *size* of a poorly designed survey will not improve its reliability. The pollsters should have used an unbiased method. Selected properly, the poll could have included only about 1,500 people and been much more reliable than the one with over 2,000,000 in the sample.

The pollsters might have used a **stratified random sample**, which takes into account various characteristics known to be relevant to the matter being polled. We know, for example, that a person’s voting behavior is related to characteristics such as gender, race, age, income, education, geographical habitation, and political party affiliation.

If the target population is known to be differentiated according to factors such as political party, income, education, occupation, age, gender, religion, geographic location, etc., then the method of selecting the sample should ensure that the relevant social strata are adequately represented. The researcher should control the study to ensure that relevant sub-groups are neither over- nor under-represented. “Thus, every important or relevant social stratum within a given population must be present in proper proportion in the sample if the latter is to be an adequate representation of the [target population]. The greater the number of important characteristics which the sample has in common with the aggregate or the group as a whole the more representative and, therefore, the more reliable the sample will be” (Werkmeister 1948: 477).

The *Literary Digest* sample could have included a representative number (relative to those who vote) of democrats, laborers, women, southerners, college graduates, farmers and any other group whose interests or characteristics would be significantly related to their vote.

5. The questions asked

The Annenberg pollsters noted that “variations in the wording and order of questions... may lead to somewhat different results.” To evaluate an opinion poll properly, you must consider what questions were asked by the pollster. The questions must be clear and they must be fair. Some polls ask the questions in different order to different people, to offset any bias that might occur because of the order of asking the questions. If you are told only the results of the poll, you cannot be sure it is unbiased, even if the method of selection was unbiased and the size of the sample is satisfactory. The questions asked might be leading or misleading questions. The questions might be asked in such a way as to suggest a particular response. The questions might be loaded, be based on a false dilemmas, make subtle appeals to authority, popularity, etc. Certain words might tend to evoke positive responses, while other words might tend to evoke negative responses.

Here are the questions asked in the Annenberg survey:

Would you favor or oppose an amendment to the U. S. Constitution saying that no state can allow two men to marry each other or two women to marry each other?

Would you favor or oppose a law in your state that would allow two men to marry each other or two women to marry each other?

These questions seem clear enough, but are they fair? They seem straightforward enough. The wording isn't obviously leading or biased. (Note: some polls allow respondents to answer "no opinion.")

Asking unbiased questions about controversial issues is very difficult. For example, what questions would you ask about *abortion*? Would you ask: *Are you for or against abortion?* What kind of question is that? Is it even meaningful? How can a person be *for* abortion? How can anyone be *against* abortion? If you are *for* abortion does that mean you do not want anyone to have a baby or that you want all pregnant women rounded up and brought in to abortion clinics? Does it mean that you are for not interfering with a woman's decision to have an abortion? If so, then that is quite a different matter from being *for* aborting all fetuses. Does it mean that you would have an abortion if you got pregnant? Does it mean that the more abortions there are the happier you are? Does it make sense for a male to say he is *for* abortion? Does it make sense to ask a pregnant woman if she is for or against having an abortion? The pregnant woman who is trying to decide whether to have an abortion would have her decision trivialized by reducing it to being *for* or *against* abortion. How could anyone be *for abortion* in general? You might say that one is either for or against having a specific abortion but not for or against abortion in general.

Would you ask: *Should abortion be illegal?* That seems like a straightforward enough question, but is it? The question assumes that all abortions are the same. Imagine asking *Should firearms be illegal?* Most of us would at least want to make a distinction between firearms used by the military and the police in their efforts to protect society, and firearms used by gangsters and thugs in their efforts to harm society. The question forces us to treat all firearms, used by anyone for any purpose, as equal. Likewise, for the question *should abortion be illegal?* The question forces us to treat all abortions, done at any time for any reason by anyone, as equal. It forces us to treat the abortion performed in order to try to save the pregnant woman's life as equal to the abortion performed in order to take advantage of a free cruise. It forces us to treat an abortion in the first days or weeks of pregnancy as equal to an abortion in the seventh or eighth month of pregnancy. It forces us to treat an abortion of a rape victim as equal to an abortion of a woman who tried to get pregnant and then changed her mind about having a baby. Some people who have done some deep thinking about abortion think these distinctions are important. However, the question forces the respondent to take an all or nothing stance.

Would you ask: *Should the decision to have an abortion be up to the pregnant woman?* This question, too, seems straightforward enough. Yet, is it suggesting that either a pregnant woman be given total control over her pregnancy or be forced to carry her fetus to full term? Is the question forcing us to focus our attention on the power struggle for control over the woman's body? If so, it is leading us to say *yes*, for most of us are trained from birth in the U.S. to value the individual and privacy against the incursions of governmental power.

Sometimes, it is obvious that a question is biased. For example, if people are asked if they favor "affirmative action" they respond more positively than if asked if they favor "preferential treatment." On the other hand, overall results differ when people are asked if they "favor" something rather than if they "are opposed to" the same thing. As noted by the Annenberg pollsters, even the order in which questions are asked can influence the responses. For example, in 1939 two polls asked whether the U.S. should allow its citizens to join the German army. In one group, 31% said "yes." In the other, only 22% said "yes." The first group was first asked *Should the U.S. permit its citizens to join the French and British armies.* The second group was first asked *Should the U.S. permit its citizens to join the German army.*

In 1996, the National Democratic Party sent out a nationwide poll, allegedly to get input for the Democratic platform. Respondents were asked, "Which position to you believe the Democrats should adopt regarding Medicare?" The choices were (A) "To make the Medicare program more financially stable, we should make deep cuts in Medicare spending—even if it means higher premiums and a lower standard of care." (B) "Some spending reductions are necessary but they should focus on controlling health care costs. The best way to fund future Medicare spending is to cancel the proposed Republican tax breaks for Americans making more than \$100,000 per year." (C) We should not cut Medicare spending. We should cancel the proposed Republican tax break for the rich and reduce spending in other areas." (D) No opinion or none of the above. The likelihood is slim that many would choose higher premiums and lower quality of care. Some might have no opinion. The majority will choose either B (some spending reductions) or C (no reductions), but in either case they must also choose to favor something that has nothing to do with Medicare: a proposed Republican tax reduction. However, the whole point of this question may have been to poison the well against the Republican proposal by characterizing it as a "tax break for the rich."

Another question in the survey asked “Do you favor or oppose Republican efforts to eliminate the 100,000 new police positions funded under the President’s 1994 Crime Bill?” How many citizens will oppose hiring new police officers? (It depends on how you ask the question! *Do you favor hiring twenty new police officers even if it means closing down the City Park & Recreation programs and the local high school?*)

Finally, even if the questions are unbiased the pollster can bias the answers by priming the subjects. For example, early in 1999, Juanita Broaddrick claimed that President Clinton sexually assaulted her 21 years ago. A CNN/Gallup/USA Today poll found that 34% believed the claim was true and 54% did not believe her. A Fox News/Opinion Dynamics poll, however, found that 54% believed her and only 23% thought her claim was not true. An editorial in the *Wall Street Journal* explained the difference as due to the fact that the CNN poll used the word ‘rape’ in its question, while Fox used ‘sexual assault.’ CNN asked “Do you think Broaddrick’s allegation (of rape) is true?” Fox asked “Based on your knowledge of Bill Clinton, are the allegations (of sexual assault) more likely to be true or not true?” On the surface, this might seem like a plausible explanation. However, the *Wall Street Journal* intentionally omitted from its editorial some crucial information. *The two polls were done at different times.* The Fox poll was done *before* an NBC *Dateline* interview with Broaddrick. The CNN poll was done *after* the interview. Before the *Dateline* interview, she had not used the word ‘rape’ to describe the incident; during the interview, the incident was repeatedly referred to as *rape*. So, the two polls used different words because those were the words Broaddrick was using at the time the polls were done. The pollsters were not trying to bias the results by using different words.

In addition to not informing its readers that one poll was done *before* the television interview and the other *after*, the *Wall Street Journal* failed to report that Fox had primed its subjects with a set-up question: “*Last week The Wall Street Journal published 20-year-old sexual assault allegations by Juanita Broaddrick against Bill Clinton. Broaddrick says while forcing her to have sex, Clinton tore her pantyhose, held her down, and bit her lips. She also says that Clinton tried to apologize 13 years later, just before announcing his campaign for president. Have you heard about Broaddrick’s allegations against Clinton?*” Compare this question with the set-up question asked by CNN: “*A woman from Arkansas named Juanita Broaddrick has recently stated that Bill Clinton raped her in 1978. Clinton has denied the allegation. Have you heard the news about this allegation before now, or not?*”

By the way, Peter Hart and Robert Teeter of the *Wall Street Journal* did their own poll *after* the editorial appeared, using the word ‘rape’ in their question, and found that 50% did not believe the allegation by Ms. Broaddrick. Thus, it seems that it was the timing and the priming done by the pollsters, not the words used in the question that led to the different results.⁴

6. Public Opinion Polls

A variant of managing or faking the news, discussed in chapter three, is the use of opinion polls to create news where none existed before the poll was done. Just as advertisers use imagery to involve the audience, TV stations and newspapers try to connect with their audiences by using gimmicks like taking an opinion poll on such issues as whether there should be abortion on demand or whether viewers are for or against a constitutional amendment to prohibit desecration of the flag or whether they think we should get out of Iraq. The polls are unscientific and their results have no meaningful news value. Yet, the results are often presented in news stories as if they were accurate measurements of the community’s pulse.

On the other hand, television stations and newspapers have marketing directors whose job is to do viewership or readership surveys. Such surveys not only determine what pretty boy or girl stays hired (on TV) but what kind of news and information will be presented to the public.⁵

Of course, any political adviser or consultant, whose job is to create news which would work in their candidate’s or boss’s favor, is very familiar with manipulating the press or the public through opinion polls. Selective release of poll data enables political managers to manipulate the media and public opinion about candidates or office holders. The news media can hardly be a reliable source

59% percent of Americans can name the Three Stooges, but only 17% can name three U.S. Supreme Court justices.--Washington Post Poll, Washington Post, October, 12 1995

of information about government and special interest blocs if the government and special interest groups themselves are managing information that might affect our opinion of them.

Public opinion analysis has been an integral part of presidential politics for the past 20 years.⁶ Robert Teeter, a pollster and adviser to George Bush, used polls to find out what issues Bush should push and which ones he should avoid or go lightly on. He focused on political issues on which the Democrats were presumed to be vulnerable, such as crime and taxes.⁷

The problem with public opinion is that it is ephemeral. It is inherently unstable and shifts with changing winds of information and events. “Change the wording of a question, and you change the opinion. Change the order of questions, and you change the opinion. Change the time of day you conduct the poll, and you change the opinion.”⁸ “Journalists, trained to seek out fact, increasingly have failed to make a distinction between objective, quantifiable fact and opinion. The result has been that ephemeral opinion has, all too often, begun to substitute for objective fact in the diet of information the media provide.”⁹ What we would like journalists to do is inform us of what is going on. What we are getting is journalists informing us of our *opinion* of what we think is going on.¹⁰

Norman Bradburn, director of the National Opinion Research Center at the University of Chicago, calls these non-scientific polls *SLOP surveys*. SLOP is an acronym for *self-selected listener opinion polls*. He compares them to radio talk shows: they attract a slice of America that is not representative of the country as a whole. “As a result, SLOP surveys litter misinformation and confusion across serious policy and political debates, virtually wherever and whenever they are used.”¹¹ The inaccuracy of such polls should be obvious. Those who call in to give their opinion are self-selected rather than randomly selected. It appears that people who are willing to call in their opinion once will sometimes call in their opinion more than once. For example, in a *USA Today* call-in poll 81 percent of the more than 6,000 respondents said that “Donald Trump symbolizes what made the U.S.A. a great country.” However, 72 percent of the favorable calls came from two telephones in one insurance company office. Another call-in poll, conducted by *Parade* magazine in August 1990, found that the “overwhelming majority of the nearly 300,000 respondents who participated ... were opposed to abortion.” (Scientifically done national surveys consistently find that a majority of Americans favor allowing abortion.) *Parade* later admitted that 21 percent of the callers “may have voiced their opinions more than once.”

CBS tried the gimmick of call-in polling in “America on the Line,” which featured two surveys conducted immediately after President Bush’s State of the Union speech. There were 314,786 self-selected callers in one survey and 1,241 adults previously selected by a more scientific method in the other survey. The latter was to act as a check on the call-in survey. CBS’s Dan Rather commented on the similarity of results in the surveys, a sentiment that was echoed the next day in the *Washington Post*, which wrote, “by and large, the two polls produced the same or similar results.” The facts, however, do not support this judgment. “On two of the nine questions asked in both polls, the results differed by more than 20 percentage points. On another five, the differences were 10 percentage points or more.”¹² Kathy Frankovic, director of surveys at CBS, said that “the important thing is that we are engaging people.” What kind of news can we expect when the purpose is *to engage people* rather than to inform them?

7. Evaluating polls: another example

Many newspaper and television stories focus on polls that are scientifically done but which provide little useful information. An Associated Press (AP) story appeared in a local newspaper that is typical of the statistical amusements often reported on by journalists.

The article had a headline that read **MANY WOMEN CONFESS THEY MARRIED WRONG MAN**. The opening line of the article read: “Only half of all women who have been married one time say they would marry the same man if they had it to do over again, according to a new survey.” We are told that the survey was based on more than 56,000 questionnaires taken from *Women’s Day* magazine. Pollsters Yankelovich, Skelly & White chose 3,009 of the forms at random and tabulated the results. “The resulting statistics have a margin of error of plus or minus 1.8 percent, according to the magazine,” said the AP story. Of course, the pollsters found many other interesting things. For example, they found that *only* 38 percent said they would *not* marry the same man again. They also found that of those married more than once, 63 percent said they *would* marry their present

husbands again; *only* 24 percent said they wouldn't. Thus, the headline does not quite fit the data. Yet, how many people would read a story with the headline MANY WOMEN SAY THEY WOULD MARRY SAME MAN AGAIN.

To begin our evaluation of the poll, we ask: *how? how many?* and *what?* How was the sample selected. *How many* were in the sample? *What questions* were asked? We are told that 3,009 questionnaires (out of 56,000) were evaluated. We are told that the sample was a random sample. We have no reason to doubt that a reputable firm, which makes its living by its reputation, would not have selected a random sample. In any case, it would be quite simple to get a random sample from this population of 56,000 questionnaires. All we need to do is devise a method that will give each respondent an equal chance of being selected for the sample. For example, we might assign a number to each questionnaire and have a computer randomly generate three thousand numbers between 1 and 56,000. The point is that we do not have to examine each of the 56,000 questionnaires in order to arrive at valid statistical generalizations about the target population. Nevertheless, when examining the *Woman's Day* poll we discover that although the method of selecting the sample is proper and the sample size is more than adequate, the poll's data is completely misused.

First, what is the population about which the conclusions of the article are drawn? The population is said to be *women*. But the population from which the sample was drawn was a group of 56,000 women who answered a questionnaire published in *Women's Day*. How likely is it that the readers of this magazine who responded to the questionnaire are typical of women in general?

Who conducted the survey? We do not know who wrote the questions, nor do we know what the questions were, for they are not mentioned in the AP article. We know that *Women's Day* published both the survey questions and the analysis of the results. We know that the analysis was done by Yankelovich, Skelly and White (a reputable polling firm). It is not possible to evaluate the questions for bias, since the newspaper article does not list them. In any case, the magazine may have done the study to find out about its readers, not about women in general. However, the newspaper article presents the results as if the survey had a more general application.

Since the population from which the sample was taken is likely to overrepresent certain kinds of women and underrepresent others, the sample is likely to be biased. This is true if the sample is used to draw conclusions about women in general. It is even doubtful that this method justifies conclusions about readers of the magazine.

Finally, note the claim that "the resulting statistics have a margin of error of plus or minus 1.8 percent, according to the magazine." This statistic seems to give the study an aura of respectability and accuracy. However, it is misleading. It only means that if they took a random sample from the questionnaires, 95 out of 100 times they would get the same results plus or minus 1.8 percent. That is, it is highly probable that they would get the same useless results—regarding women in general—no matter how many times they repeated the sampling.

8. The Nielsen ratings

The A.C. Nielsen company (Nielsen Media Research) probably wields more power over television programming than any other entity. The basis for their power is the acceptance of their rating system, a system of surveying television viewers and estimating how many people watch particular programs. Many advertisers agree to base their rates on the Nielsen data. Since television programs exist mainly to provide advertising, a Nielsen rating can make or break a program.

A.C. Nielsen uses several methods of surveying television viewers. Unlike most companies that do polling and surveying of opinions and behaviors, Nielsen does not rely on telephone surveying. They do about two million questionnaires a year by mail in the country's 211 media markets. Prospective participants are asked to keep a diary of their TV viewing for the week. This is done four times a year during what is known as "sweeps week." However, their most well known method involves attaching a box to the television sets of those selected to be Nielsen participants. This box monitors the set, letting the Nielsen folks know when the set is on and what

There is an average of 50 ads for "junk food" during Saturday morning cartoon shows.--Center for Science in the Public Interest report, Washington DC, 1991.

The average child views some 30,000 commercials a year. --Consumer Reports, February 1998.

channel it is tuned to. They even monitor the recording of programs on VCRs. Nielsen also provides about 5,000 homes with a “people meter,” an electronic pad for data input by anyone who watches TV in the home. Each person in the home is assigned a button on the pad, which they are supposed to punch whenever they are watching TV. Even visitors have a button to punch. Participants are asked to punch out whenever they are leaving the TV area, even if only to go the kitchen for a snack or to the bathroom. If the TV is tuned to the same channel for 70 uninterrupted minutes, a light flashes on the meter. If a viewer does not push the “yes” button (to indicate he or she is still watching TV), the Nielsen box stops registering TV input.

The Nielsen box is connected to a telephone line. During the middle of the night it uploads data to Nielsen’s computers in Dunedin, Florida. About twelve hours later, the data has been analyzed and is ready for publication, though Nielsen press reports are generally issued weekly.

The 5,000 homes with the box and the meter agree to participate for two years. Nielsen researchers select a stratified random sample of households from across the nation. New participants are added at a rate of several hundred a month, as current participant’s terms expire. The first week of data from the home is ignored, to give the family time to adjust to the setup. Nielsen also hooks up about 500 homes for five-year periods. These homes do not have the “people meter” and they are selected from the 36 largest media markets, rather than from across the country. Any participant who publicly acknowledges that they are a Nielsen home is disconnected.

The American Psychological Association estimates that the average American child will view 8,000 murders and 100,000 other acts of violence on television before finishing elementary school.

One Nielsen rating point is the equivalent of some 970,000 households, i.e., about 1% of all households, or about two million viewers. The Nielsen people provide a number of statistics regarding our television habits. For example, in 1988 they reported that the average American watches 3 hours and 46 minutes of TV each day. That translates into more than 52 days of nonstop TV-watching per year. By age 65 the average American will have spent nearly 9 years watching TV. The television is on for an average of 7 hours and 12 minutes a day in the typical American home. That’s a lot of electricity. Two-thirds of us watch TV while eating dinner. That’s a lot of indigestion. The final episode of NBC’s 11-year hit sitcom “Cheers” in May, 1993, attracted 93.1 million viewers (64% of all viewers), with a 45.5 Nielsen rating. Ten years earlier, the last episode of M*A*S*H had a rating of 60.2 (some 77% of all TV viewers were tuned in).

Primarily, however, the Nielsen ratings are used to decide how much to charge advertisers. Over a year, a difference of one rating point could mean as much as \$100,000,000 in advertising revenue. Television programs are cancelled because their Nielsen ratings are too low. How accurate are the ratings? It is difficult to know, since there is no standard to measure the ratings against. However, some things we do know. For instance, we know that Nielsen does not measure TV viewing in hospitals, hotels, bars, college dorms, airports, prisons, and other public places. There is no way to monitor how diligent the participants are in using the people meters, and no way to know how honest they are in filling out the diaries. Equipment failures, which apparently are significant, affect the ratings. Many of the complaints about Nielsen accuracy come from the major networks, which have seen their ratings decline steadily with the advent of cable television. There have been two important responses by the major networks ABC, NBC, and CBS: (1) they have bought or started their own cable networks and (2) they are funding a rating system of their own. The system called SMART (for Systems for Measuring and Reporting Television) is run by Statistical Research, Inc. It may offer competition to Nielsen in the future, but it is being criticized as a system designed mainly to increase network ratings.

9. Statistics and tricks with numbers

When evaluating statistical studies it is important to know who did the study, how it was conducted, the size of the sample, and the questions asked of those polled. Many polls and surveys are done by self-interested groups or individuals; they’re self-serving and rarely do they inform the public of how they were conducted or how many individuals were studied. I am referring to polls taken by political candidates and their organizations, and polls taken by commercial companies and the advertising firms they hire to sell their products. Such polls and surveys

generally have little news value. The fact that you are not told how a study was conducted and how many were in the study is a good sign that the study is not trustworthy.

It is easy to be misled by statistical studies. Precise numbers make conclusions seem accurate. Sometimes, however, their very precision should make us skeptical. For example, a magazine article reported that “a record 577,357 criminals now crowd the nation’s jails.”¹³ Since the number of jail inmates changes continuously, claiming to know the exact number of inmates is ridiculous.

Numerical data can be presented in misleading ways, too. For example, if you were told that 60 percent of those polled favored a particular consumer product, you might be impressed. However, if the sample included only five members of a very large population, the large percentage is meaningless.

Statistics do not always tell the whole story, either. If evidence is omitted, a story can be misleading. For example, the *Washington Post* published a story by reporter Cheryl Thompson on March 10, 1999 in which she gave some statistics about a federal program that detained those accused of crimes in pretrial halfway houses. She noted that a number of those detained had escaped and that “at least 83 of those 226 pretrial inmates who absconded—some more than once—were rearrested on new charges, including manslaughter and armed robbery, according to District [of Columbia] and court records.” She was correct, but she did not mention that 63 of the 83 when rearrested were charged only with escaping from the halfway house. Thus, she gave the impression that the escapees had gone on a crime spree. Only one escapee was charged with manslaughter and one with armed robbery.¹⁴

Insurance companies often cite studies that show that more deadly accidents occur among young drivers (age 16-25) than among any other age group. They charge young people a correspondingly higher rate for insurance. Many times such studies fail to consider the number of drivers in the various groups or the number of miles driven on average by members of the various age groups (to establish that the young drivers are not disproportionately represented). Another example is the seemingly annual article on the high cost of college. Writers project that by the year two thousand it will cost \$25,000 a year to go to college, but they omit that they are assuming that inflation will continue at a certain rate and they fail to note that the value of those dollars in the next century won’t be quite the same as their current value. Such articles also tend to omit consideration of costs that a student would have even if he or she did not go to college, such as the cost of food and housing.

Finally, a word of caution about the term *average*—a very frequent term in statistical studies—ought to be given. There are several different meanings of the term ‘average’, and it is possible to be misled by a statistical study which uses ‘average’ ambiguously. One meaning of ‘average’ is the *arithmetic mean*. This is the figure one gets when a set of numbers are added up and divided by the number of items in the set, e.g., finding the class average on a test by adding up the scores of each student and dividing the sum by the number of students who took the test. The arithmetic mean is a good measure of the average if there are no extremely low or high values that would skew the results.

Another way of determining the average is to determine the *median*. To compute the median, a list of values must be composed in either increasing or decreasing order. If the number of values is odd, then the median is the value that falls in the middle of the list. If the number of values is even, then the median is the arithmetic mean of the two middle values. Thus, if we wanted to find the median of (70, 75, 76, 85, 90) we would find the middle value, namely, 76. If the list were (70, 75, 76, 80, 85, 90), then the median would be 78, i.e., the arithmetic mean of 76 and 80. The median represents the average of the majority of values in the list. It is especially useful if there are extremely low or high values in the list that would skew the results. A person buying a house in a neighborhood might want to know what the median cost of a house in that neighborhood is.

Another kind of average is known as the *mode*. The mode represents the value that occurs most frequently in a set of values. For example, a shoe store owner might say that the average size shoe he sells is 9½, meaning that he sells more shoes of that size than any other size. It is important that one know which ‘average’ is being referred to in a statistical study in order to properly evaluate it.

Exercise 7-1 Self-test: true or false? (Check your answers in Answers to Selected Exercises.)

1. Justified inductive generalizations can provide us with a useful means of facing the future because they are *predictive*.
2. Telephone sampling is the most common method of conducting polls for the mass media.
3. Very large samples are always representative samples.
4. A *stratified random sample* is one that randomly selects strata of society to poll.
5. Samples that are self-selected are rarely biased.
6. If the sample is *too small*, it is likely to be biased.
7. Even if the questions are unbiased, the pollster can bias the answers by priming the subjects
8. A representative sample is one that will give us a good cross section of typical items in the target population.
9. The *margin of sampling error* in a poll is a measure of the poll's accuracy and representativeness.
10. Someone standing on an assembly line, whimsically selecting items to be tested, would be collecting a random sample.
11. If a sample is a truly representative sample, then the larger the sample the more likely it is representative of the target population.
12. The term *target population* refers to the whole aggregate from which a sample is taken.
13. The margin of sampling error is a measure of errors due to biased data, erroneous calculations or inappropriate application of the statistical data.
14. The mode represents the value that occurs most frequently in a set of values.
15. A *random sample* is one selected in such a way as to give each item in the target population an equal chance of being selected.
16. A small margin of error indicates that if the sample is biased it's not because the size of the sample is too small.
17. Opinion polls are never used to manage the news.
18. Call-in polls are unscientific and although they have no meaningful news value, they are often presented in news stories as if they were accurate measurements of the community's pulse.
19. According to Kathy Frankovic, director of surveys at CBS, the important thing about call-in polls is that they engage people not that they provide any useful or reliable information.

Exercise 7-2

Discuss the following questions asked by a pollster.

1. *Do you favor retaining loud-mouthed, pro-terrorist, racial agitator Andrew Young as U.S. ambassador to the United Nations?*
2. *Do you favor stronger punishments for criminals or do you favor the current lenient system of punishment?*
3. *What is your political affiliation? (Assume the answer is 'Democrat'.) Which of the currently running democratic nominees do you favor?*
4. *Do you support your constitutional right to own and bear arms?*
5. *Do you agree with most American women that abortion should be a woman's choice?*
6. *Do you agree with the Supreme Court that women should be allowed to murder their babies at will?*
7. *Are you for or against capital punishment?*
8. *Do you support American intervention in Kosovo even if it means many American soldiers will be killed?*
9. *Do you favor releasing dangerous criminals from prison, even if they have served their full sentence?*
10. *Have you ever experienced a period of time of an hour or more, in which you were apparently lost, but you could not remember why, or where you had been?*

Exercise 7-3

There are 3,000 guests at a party. One hundred are interviewed at random and it is discovered that 70 are registered Democrats, 20 are Republicans and 10 are Independents. The margin of error is $\pm 12\%$. Discuss each of the following statements as if they were independently drawn conclusions.

1. It is highly probable that the majority of guests are democrats.
2. 70 percent of the guests are democrats.
3. It is probable that at least 60 percent of the guests are democrats.
4. It is highly probable that about 70 percent of the guests are democrats.
5. It is highly likely that the Republicans are in a minority at this party.
6. It is highly probable that most people are democrats.

Exercise 7-4

Evaluate the following arguments. [Exercises with an * are answered in Answers to Selected Exercises.]

- *1. The democrats will win the election by a landslide, according to a survey taken today at Farmer's Market in downtown San Francisco. Two hundred people were interviewed during the lunch hour and 75 percent said they plan to vote democratic in the upcoming state and federal elections.
2. To find out what percentage of their income a family of four spends on food, a pollster goes through the phone book of a major city and randomly selects 1,500 different names to call. Three hundred and twenty five of those called identify themselves as families of four. Three-fourths of these say they spend at least 30 percent of their income on food. The pollster concludes that in the United States 75 percent of the people spend 30 percent of their income on food.
3. "The more sexual activity married women have, the more highly they think of themselves--a phenomenon that does not hold true for men. These are two of the many conclusions from a small study of 11 couples in marriage counseling and 11 couples who responded to an ad requesting 'happily married' couples to participate in a research project." *Psychology Today*, February, 1980
4. "People on welfare are a bunch of lazy cheats--parasites living off the labor of us hardworking folks. Last week I saw this lady at the grocery store. She was all dressed up in a nice outfit, wearing nice rings and an expensive watch. She bought steaks and expensive roasts, while all I could afford was chicken! And what did she pay for her food with? Food stamps!"
- * 5. "Men are basically insecure babies. I know; I was married to one for twenty years."
6. The majority of working class Americans supports the government's import quotas on foreign automobiles, as is evidenced by a recent poll taken outside of General Motor's plant in nearby Dearborn. More than 90 percent of the 47 General Motor's employees polled believe that "it is fully consistent with capitalism's ideal of free enterprise" to restrict the importation of foreign automobiles in order to help bolster sales of American-made cars.

Exercise 7-5

The following consists of a base argument. Look at the base argument and get a general notion of the sufficiency of the premises for the stated conclusion. Following each base argument is a list of statements. The statements either change the strength with which the conclusion is asserted to follow from the evidence or they offer alternative premises. Treating each additional statement separately, determine whether it strengthens, weakens or has no effect on the argument. Give reasons for your views.

BASE ARGUMENT: A newspaper reports that a recent study showed that persons who were registered Democrats were 75 percent more likely to vote for a Democratic candidate than for a Republican candidate. The study was a telephone survey of six hundred registered Democrats. The margin of error was $\pm 5\%$.

- a. The survey was done by a firm known to have done business with Democrats.
- b. Only San Diego was included in the study.
- c. The study was done in such a way as to insure that the sample included representatives of several geographical areas of the country.
- d. All of those in the study were women.
- e. Those conducting the study are all married.

h. The study sample is a good cross-section of persons, i.e., is a stratified random sample.

Exercise 7-6

Find two or three articles based on polls in newspapers or magazines and evaluate them. Comment on the method of selecting the sample, the size of the sample, the fairness of the questions asked and the margin of error. Is the sample likely to be representative? Why or why not? Is there anything not mentioned in the article about the study which should have been mentioned?

10. Analogical reasoning

Sampling involves studying a part of a population (the sample) and drawing a conclusion about the whole population. It is both *empirical* and *predictive*. The justification of sampling depends on the representativeness of the sample, which depends on how the sample was selected and how many were selected.

Sampling is a type of **simple induction**. When we generalize from several particular items or experiences, we are doing in a non-systematic way what the pollster is doing. For example, when we conclude that “General Motors makes great trucks” based upon a few experiences with their trucks, we are using simple induction. If the items we have experienced are typical and we have experienced a sufficient number of them, then our conclusion is a sound one.

Another type of inductive reasoning we commonly engage in is **analogical reasoning**, which involves comparing one thing or group of things with similar things or groups. Like sampling, analogical reasoning is both empirical and predictive. However, *the justification of analogical reasoning depends on the relevant similarities outweighing the relevant dissimilarities of the items being compared*. If the relevant similarities are more significant and more numerous than the dissimilarities, then the analogy is a sound one.

The basic structure of analogical reasoning is as follows:

1. **X and Y have strong relevant and significant similarities.**
2. **If X & Y have strong relevant and significant similarities, then what is true of X is probably true of Y and vice versa.**
3. **‘a’ is known to be true of X.**
4. **So, ‘a’ is probably true of Y.**

The letter ‘a’ represents some quality or relation, such as *earns over \$50,000 a year*. ‘X’ and ‘Y’ may represent an individual—such as *Jane Woe*—or some individuals—such as *40 percent of the adults in Woeville*—or all individuals of the same type—such as *all the adults in Woeville*.

Below are three examples of analogical reasoning.

- 1. Humans and rodents have strong, relevant & significant similarities in their reproductive systems.**
- 2. If humans and rodents have strong, relevant & significant similarities in their reproductive systems, then what is true of human reproduction is probably true of rodent reproduction and vice versa.**
- 3. Human females sometimes produce an enzyme that prevents spermatozoa from penetrating her eggs.**
- 4. So, it is probably true that female rodents sometimes produce an enzyme that prevents spermatozoa from penetrating her eggs.**

Example 1 – analogical reasoning

- 1. The philosophy class I am taking this semester has strong, relevant and significant similarities to the philosophy class I am enrolled in for next semester.**
- 2. If philosophy classes have strong, relevant and significant similarities, then what is true of one of the classes is probably true of the other.**
- 3. I really enjoyed the philosophy class I am taking this semester.**
- 4. So, I probably will enjoy the philosophy class I will take next semester.**

Example 2- analogical reasoning

- 1. Double-celling of criminals has strong, relevant and significant similarities to forcing children to sleep in bunk beds.**
- 2. If double-celling of criminals has strong, relevant and significant similarities to forcing children to sleep in bunk beds, then what is true of double-celling is probably true of double-bunking.**
- 3. Double-celling of criminals is cruel.**
- 4. So, forcing children to sleep in bunk beds is probably cruel.**

Example 3 – analogical reasoning

Justifiable analogical arguments must be based on *strong* analogies. That is, it must be true that the relevant and significant similarities outweigh the relevant dissimilarities. In addition, as with sampling, we must make sure that the conclusion is stated with the appropriate strength. The stronger the analogy, the stronger the conclusion should be.

To see the difference between simple induction and analogical induction, compare the reasoning of **Example 2** above with non-analogical reasoning to the same conclusion. Suppose that a person took a philosophy class and found it to be extremely rewarding. In planning her schedule for the next semester she decides to take another philosophy course, fully expecting it to be rewarding also. We might reconstruct her reasoning as follows:

1. **The philosophy course I took was extremely rewarding.**
2. **So, all philosophy courses are extremely rewarding.**
3. **Therefore, the philosophy course I take next semester will be extremely rewarding.**

Example 4 – simple induction to a hasty generalization and a valid deduction

If the student were to reason this way, we would find fault in the process of drawing a conclusion about *all* philosophy courses based on a single experience. She would be making a **hasty generalization**. There are too many variables to justify a universal generalization about the “rewardingness” of philosophy courses based on one person’s single experience.

Now compare the reasoning in **Example 4** with that of **Example 2**. In the case of **Example 2**, no generalization is made about philosophy courses. Instead, an *analogy* is drawn between the course already taken and the course offered next semester. The adequacy of the reasoning depends on the relevant similarities between the two courses outweighing the relevant dissimilarities between them.

Generally speaking, the greater the number of characteristics that different items or types of items share, the greater the likelihood those items or types of items will share some further characteristic. Also, generally speaking, the fewer the characteristics that items or different types of items share, the less likely it is that they will share some further characteristic. However, the *number* of shared characteristics is not nearly as important as their *relevance* and *significance* to the issue at hand. In **Example 2**, the issue is whether the philosophy course the student enrolled in for next semester will be rewarding. What characteristics of a philosophy course are relevant to its being rewarding or not? Does the proposed course share in common many of these characteristics with the course already taken? Are any of these characteristics known *not* to belong to the proposed course?

We will assume that the following list of characteristics is made up by the student in response to the first question.

Characteristics relevant to a philosophy course being rewarding

<p>Course content. The instructor. The text book(s). Interest in the subject.</p>	<p>Course requirements. Where the course is to be taught. Hour at which the course is offered.</p>
--	---

If the instructor for the two courses is the same, that strengthens the argument. If the instructors are different persons but very similar in their styles and personalities, that would also strengthen the argument, but not quite as much as if the instructor were the same for each course. If the instructors were not only different, but also very different in personality and teaching style, that would weaken the argument.

What about the course content? If the course taken was in the history of philosophy and the proposed course is in mathematical logic, would that strengthen or weaken the argument? What if both courses were in the history of ancient philosophy? What if the text book in both courses was identical (different chapters being read in each course)? What if the books were different?

What if the requirements for the course taken were all in the form of take-home essays and papers, but the proposed course will have only in-class essays and/or true/false questions on exams?

In short, relevant similarities strengthen the argument; relevant dissimilarities weaken the argument. Whether or not the student reasoned well in the example depends on the strength of the relevant similarities outweighing the strength of any significant dissimilarity.

Finally, many characteristics that are similar or dissimilar about the two courses are *irrelevant*. For example, each course will require the same registration fee, the classrooms will have doors, the teachers will be short, tall, blue-eyed, and so on, but these are irrelevant. Thus, you need not consider them.

11. False analogies

Example 3 reveals an interesting characteristic of many analogies. Two *concepts* are compared: *double-celling of criminals* and *forcing children to sleep in bunk beds*. An argumentative analogy is a good one if the relevant similarities of the items being compared outweigh the relevant dissimilarities, which is not the case in this argument. While it is true that there are some similarities to forcing a criminal to share a cell with another criminal and forcing children to sleep in the same room in bunk beds, there are so many relevant and significant dissimilarities between the two that any argument based on an analogy between the two will be a **false analogy**. False analogies are often seductive, and a critical thinker must remember to consider relevant *dissimilarities* not likely to be brought up by the one making the argument.

Here is another example of a *false analogy*:

Since people are like vines, they will grow up unhealthy, unruly and wild, if they are not disciplined and cared for when they are very young.

Example 5 – a false analogy

There are too many relevant and significant dissimilarities between vines (or any other kind of plant) and human beings with respect to the issue of growth for this analogy to be relevant. Similarities by themselves are not sufficient to prove any conclusion in an analogical argument. Thus, while it is true that raising vines and raising children share some things in common, the relevant and significant differences between the two outweigh any similarities they might have. If this were not so, then it would make perfect sense to buy a book on raising vines to help you figure out the best way to raise your child. Likewise, it would be wise to buy a book on child care before starting your garden.

12. Analogies in illustrations

Not all analogies are argumentative. Many analogies compare very dissimilar things in order to try to *exemplify* or *illustrate* a point. For example,

People are like vines. If they are not disciplined and cared for when they are very young, they will grow up unhealthy, unruly and wild.

Example 6 - an illustrative analogy

In **Example 6**, the analogy tries to *illustrate* a point. This is a perfectly good use of an analogy. **Example 6** uses the same analogy as **Example 5**; however, in **Example 5** the analogy tries to *prove* a point. Since their function is different, the analogies must be evaluated differently. An illustrative analogy is a good one if it clarifies or helps makes a point more vivid. An argumentative analogy is a good one if its premises adequately support its conclusion, i.e., if the relevant and significant similarities outweigh the dissimilarities.

13. A famous analogy: argument or illustration?

One of the most famous analogies in the history of law is the analogy used by Oliver Wendell Holmes in *Schenk v. United States*. Holmes compared passing out anti-draft leaflets during war time and falsely yelling “Fire!” in a crowded theater. The analogy, however, was *not* an *argumentative* analogy. It was an *illustrative* analogy. Holmes was trying to illustrate his point that passing out anti-draft leaflets during war time was a “clear and present danger” to the national interest and was no more protected by the First Amendment than “falsely shouting fire in a theater and causing a panic.” Attorney Allen Dershowitz calls it an “inapt analogy” because the

example of [falsely] shouting ‘Fire!’ obviously bore little relationship to the facts of the Schenk case. The Schenk pamphlet contained a substantive political message. It urged its draftee readers to *think* about the message and then--if they so chose--to act on it in a lawful and nonviolent way. The man who [falsely] shouts ‘Fire!’ in a crowded theater is neither sending a political message nor inviting his listener to think about what he has said and decide what to do in a rational, calculated manner. On the contrary, the message is designed to force action *without* contemplation.”¹⁵

Holmes was not trying to *prove* that passing out anti-draft leaflets during war time is not protected by the First Amendment *because* such activity is just like falsely yelling “Fire!” in a crowded theater. He was not saying that the two examples are so similar that since the one is not protected speech neither should the other be. Holmes was trying to illustrate what most of us would probably accept as a truism: whether speech is protected by the First Amendment depends on the circumstances. Nevertheless, his analogy is inapt because it did not do what *illustrative* analogies are supposed to do. It did not clarify the issue. In fact, Holmes muddled up the point so badly, says Dershowitz, that most of the frequent references to it since have made it into “little more than a caricature of logical argumentation.” One of the most frequent uses of the analogy, says Dershowitz, is in arguments against censorship. The Rev. Jerry Falwell, for example, argued

Just as no person may scream ‘Fire!’ in a crowded theater when there is no fire, and find cover under the First Amendment, likewise, no sleazy merchant like Larry Flynt [publisher of *Hustler* magazine which had been sued by Falwell] should be able use the First Amendment as an excuse for maliciously and dishonestly attacking public figures, as he has so often done.

Even judges have cited the Holmes’ analogy to justify restricting pornography, picketing, and marching. Dershowitz cites numerous other examples of misuse of the Holmes’ analogy. He agrees that there are exceptions to the First Amendment’s exhortation that the “Congress shall make no law...abridging freedom of speech, or of the press.” However, Dershowitz denies that any of the exceptions bear any meaningful resemblance to falsely shouting “Fire!” in a crowded theater.

14. Analogical arguments in law

One field where analogical arguments abound is law. Judges and lawyers often defend their positions by appealing to precedents analogous to the case at hand. For example, In *Stanley v. Georgia* the court wrote, “If the First Amendment means anything, it means that a State has no business telling a man, sitting alone in his house, what books he may read or what films he may watch.” Several years later, a lawyer cited *Stanley* and argued that viewing pornography at home was analogous to committing sodomy at home. His client had been charged with violating the Georgia sodomy statute by committing that act with another adult male in the privacy of his own bedroom (*Bowers v. Hardwick*, 478 U.S., 106 S. Ct. 2841, 1986). The attorney argued that since the Court recognizes that States have no business in telling a person what to read in the privacy of his own home, they have no business in telling a person what sex acts he can perform in the privacy of his own home. However, Justice Byron White noted some significant differences in the two cases. White argued that just because the two cases involve acts done in private isn’t sufficient to make them so analogous that a State has no business in telling a person what sex acts they can perform in privacy. Reading books is a First Amendment issue (freedom of speech, freedom of the press); having sex is not. White also pointed out that there are many other acts which, though done in the privacy of one’s own home, does not protect them from State legislation: “...the possession and use of illegal drugs do not escape the law where they are committed at home.” In other words, there is a strong disanalogy between the two cases.¹⁶

An even more telling example is the analogical reasoning used by the court to decide *People v. Duglash*, 41 N.Y.2d 725 (1977). The court cited *United States v. Thomas*, 13 U.S.C.M.A. 278 (1962).¹⁷ Duglash was charged with attempted murder. The person Duglash shot, however, was already dead when Duglash fired his pistol. Since it is not logically possible to murder a dead person, it seemed plausible that Duglash could not have attempted to murder that person. To do so would be to attempt to commit a crime that is factually impossible to commit. The court ruled that the Duglash case was analogous to an earlier case, the Thomas case. Thomas had been charged with attempted rape but it was established that his victim had died before he had sex with her. The court ruled that since the defendant *believed* his victim was alive, he could be charged with *attempted* rape. Thus, the court ruled, since Duglash *believed* his victim was alive, he could be charged with *attempted* murder. It had been established in the Thomas case that all that is necessary to *attempt* to commit a crime is the *belief* that you are doing so. If your actions demonstrate that you *intended* to kill the victim, you can be said to have *attempted* to commit murder, even if the victim is dead before your action, thereby making it factually impossible to commit the crime you intend.

It is perhaps worth noting that analogical reasoning by precedent is based upon non-analogical reasoning. For example, the Thomas case was decided without precedent. How was it decided? By analysis of the concepts of ‘attempted act,’ ‘belief,’ ‘intent,’ and so on. It seems reasonable to think that even if Thomas had never existed, the same reasoning that was valid in that case would be valid in the Duglash case. Thus, it seems that argument by precedent, which plays such an immense role in legal reasoning in the United States, is ultimately a search for *authoritative support* of one’s position. However, deciding similar cases in similar ways is a requirement of consistency and fairness. Reliance on precedent also reminds us that judges must give reasons for their opinions; and those reasons must not be personal, whimsical, or arbitrary.

15. Analogical arguments in philosophy

Philosophers are also fond of analogical arguments. One of the most celebrated arguments from analogy in the history of philosophy is William Paley’s “analogy of the watch” argument for the existence of God. Paley (1743-1805), the Archdeacon of Carlisle, writes in his *Natural Theology* (1802):

In crossing a heath, suppose I pitched my foot against a *stone* and were asked how the stone came to be there, I might possibly answer that for anything I knew to the contrary it had lain there forever; nor would it, perhaps, be very easy to show the absurdity of this answer. But suppose I had found a *watch*

upon the ground, and it should be inquired how the watch happened to be in that place, I should hardly think of the answer which I had before given, that for anything I knew the watch might have always been there.

The reason, he says, that he couldn't conceive of the watch having been there forever is because it is evident that the parts of the watch were put together for a *purpose*. It is inevitable that "the watch must have had a maker," whereas the stone apparently has no purpose revealed by the complex arrangement of its parts.

One could, of course, attack Paley's argument at this point and say, as Clarence Darrow did, that some stones would be just as puzzling as a watch; for, they are complex and could easily have been designed by someone for some purpose we are unaware of. It may well be that "on close inspection and careful study the stone...is just as marvelous as the watch."¹⁸ Be that as it may, Paley's point was not that watches are inherently more interesting than stones. His point was that a watch could be seen to be analogous with the creation of the universe. The design of the watch implies an intelligent designer. This fact, says Paley, would not be diminished even if we discovered that the watch before us was the offspring (no pun intended) of another watch. "No one," he says, "can rationally believe that the insensible, inanimate watch, from which the watch before us issued, was the proper cause of the mechanism we so much admire in it--could be truly said to have constructed the instrument, disposed its parts, assigned their office, determined their order, action, and mutual dependency, combined their several motions into one result, and that also a result connected with the utilities of other beings."

Paley then goes on to claim that "every manifestation of design which existed in the watch, exists in the works of nature, with the difference on the side of nature of being greater and more, and that in a degree which exceeds all computation." The implication is that the works of nature must have had a designer of supreme intelligence who contrived to put together the magnificent mechanism of the universe. According to Darrow, this 'implication' is actually an *assumption*.

To say that a certain scheme or process shows order or system, one must have some norm or pattern by which to determine whether the matter concerned shows any design or order. We have a norm, a pattern, and that is the universe itself, from which we fashion our ideas. We have observed this universe and its operation and we call it order. To say that the universe is patterned on order is to say that the universe is patterned on the universe. It can mean nothing else.¹⁹

The problem with Paley's analogy is that the belief that the universe shows orderliness and purpose is an *assumption*. One quality of a good analogical argument is that the characteristics cited as shared characteristics must be truly shared characteristics. If there is doubt that one item (the universe) has a most significant characteristic (being designed and purposive), then the analogical argument is not a sound one.

Another philosopher, David Hume (1711-1776), took up the design analogy a few years before Paley, in his *Dialogues Concerning Natural Religion*. One of the characters in Hume's dialogue, Philo, suggests that "If the universe bears a greater likeness to animal bodies and to vegetables than to the works of human art, it is more probable that its cause resembles the cause of the former than that of the latter, and its origin ought rather to be ascribed to generation or vegetation than to reason or design" (Book VII). "The world," says Philo, "plainly resembles more an animal or a vegetable than it does a watch or knitting-loom. Its cause, therefore, it is more probable, resembles the cause of the former. The cause of the former is generation or vegetation. The cause, therefore, of the world we may infer to be something similar or analogous to generation or vegetation."

Exercise 7-7 Self-test: true or false?

- 1 In an analogical argument, generally speaking the greater the number of relevant similarities between items being compared, the greater the probability they will share some further related characteristic.
- 2 Some analogies are made to illustrate, not prove, a point.
- 3 In analogical reasoning, the *number* of shared characteristics between items which are compared is more important than the *relevance* of the characteristics to a given conclusion.

- 4 A false analogy is an analogy used to prove a point when there are no relevant differences between the items being compared.
- 5 Since there are many dissimilarities between humans and other mammals such as laboratory rats, any conclusion drawn for humans based on studies of rats is unsound and improbable.
- 6 Analogical arguments are rarely used in law.
- 7 Common sense is all that is necessary to detect whether or not a characteristic is relevant in an analogical argument.

Exercise 7-8

Analyze the following analogical arguments according to the guidelines discussed in this section: compare the relevance, strength, and significance of the similarities with the dissimilarities, and determine whether the conclusion is stated with the appropriate strength or is too strong.

- *1. "...Among these countless planets there are conditions of heat and light equivalent to those we experience on earth; and if this is the case, and the planet is near our age and size, there may very likely exist human beings like ourselves, probably with strange costumes and still stranger manners." (Robert Goddard, "The Habitability of Other Worlds," unpublished paper cited in *Broca's Brain* by Carl Sagan (New York: Random House, 1979), p. 225n.)
2. Since eating only one kind of food would drive you to hate that food, it follows that having only one wife will lead you to hate that person eventually.
3. "The one and only film I saw which was directed by Wertmuller disgusted me; so, her new film will probably disgust me, too."
4. It is absolutely necessary to force atheists to accept our religious beliefs. A reasonable person would not think twice about using force to prevent a deluded person from jumping off the edge of a steep cliff. Force is sometimes necessary when dealing with children and barbarians.
5. I wouldn't give that rotten food to my dog. If it would make me sick, it will probably make my dog sick, too.
6. "I've been buying NEW IMPROVED DRUDGERY laundry detergent for years. Each box of detergent has been great, so I expect the next box I buy will be great, too."
7. People should live together before getting married because you shouldn't buy a pair of shoes without trying them on.
8. I don't think we should drink from this water hole, Mo. There are an awful lot of horse and cow carcasses lying at the water's edge. I think it's poisoned, and if it can kill a horse it can probably kill us, too.
9. Crime is like a cancer in the organism of the body politic. And like a cancer, unless it is completely excised, it will spread and eventually destroy the whole organism. The criminal, therefore, if allowed to live amongst the rest of us, will eventually spread his or her disease throughout the entire society, and we will all perish.
- *10. A lawyer for the Becton Regional High School board of trustees in East Rutherford, N.J., defended a school plan to require all students to submit to a blood and urine test administered by the school nurse or doctor, or, if the students choose to do so, by a family physician. The tests would be used to determine drug usage. The drug testing, said the lawyer, "would be no different from screening for communicable diseases, such as tuberculosis or lice."
11. It is probable that there are other minds and that I am not the only person with a mind. I don't know this for certain, but I see many other creatures who look like me, talk like me, act like me, who give signs of having a mind. That is, I know that when I talk as they do and act as they do it is because I have a mind capable of thought and intelligent discourse and action. So, I conclude that these bodies I observe driving cars and making change at the grocery store are most likely persons with minds and not robots or androids.
12. 'As President, I am the captain of the Ship of State. And, as on a ship it is *mutiny* to disobey the captain's order, so too in matters of the common good, when your President gives an order, to disobey is mutiny, that is to say *treason!*'
- *13. President Harry Truman once argued something to the effect that the United States never should have stopped atmospheric testing of nuclear bombs. 'Where would the world be today,' reasoned Truman, 'if Thomas Edison had stopped doing his experiments with the electric light bulb?'
14. (Letter to the editor:) "In all of the news media the litany of doom is prevalent. Churches and schools are full of it: 'If there is atomic war everything ends, nothing survives.' I would almost swear that there are thousands of people going about their business in Hiroshima and Nagasaki."
15. (Letter to the editor:) My neighbor leaves his doors and windows unlocked; also, he considers a burglar alarm too costly. Is he not careless? My doors and windows are locked; I have installed a burglar alarm. Furthermore, I have a gun to use in the event an intruder manages, somehow, to attack me. Which house is more likely to be entered? Does this mean I want a confrontation? ...President Reagan has no desire for war. He is only 'protecting the house.' Nuclear war is out. I refuse to believe the whole world wants to commit suicide."
16. "[Educated people] bought their education just like I buy my land and I pay taxes on the land, but they don't have to pay taxes on their knowledge and that irks me....I use the land to earn my living and they use their knowledge to earn theirs. If the

government can tax my investment, then it should tax their knowledge.” (Dairyman Earl E. Chapman, cited in Robert Baum, *Logic*, 2nd ed., (New York: Holt, Rinehart & Winston; 1980), p. 431.

*17. (Letter to the editor:) “Superior Court Judge Beverly B. Savitt has proclaimed that the double-celling of criminals is cruel and unusual punishment. Does this mean that children who are forced to sleep in bunk beds because their tax-burdened parents cannot afford larger houses can now sue their parents for cruel and unusual punishment?”

18. In a debate on spankings in school, Paul V. Armstrong, president-elect of the West Virginia Association of Elementary School Principals, said that with proper controls, corporal punishment can be an effective disciplinary tool. “I do not believe administering corporal punishment for fighting teaches aggression any more than I believe receiving a speeding ticket teaches you to be a race car driver,” Armstrong told a U.S. Senate Judiciary subcommittee on juvenile justice.

19. “All the conspicuous features on the surface of the moon are the result of impacts. These features include not only the craters, which plainly advertise their origin, but also the great maria, or “seas,” which are craters that filled with lava following the impact of very massive objects. Most of the impacts took place during a relatively brief period about four billion years ago, when debris left over from the formation of the solar system was swept up by the planets and their satellites. The earth probably received as heavy a pelting as the moon did, and it therefore must have been densely cratered. (“Science and Citizen, *Scientific American* June 1976.)

20. Several experiments have shown that the life span of mice can be significantly increased by a diet which includes certain toxic substances in small amounts. It is therefore highly probable that human beings could significantly increase their life spans by following a diet which includes small amounts of toxic substances.

21. “We judge a man’s intelligence by his words and deeds. The core of the argument for artificial intelligence lies in the fact that computer systems can be observed, by their words and deeds, to behave in similar ways. A machine can learn. It can respond (when programmed with sufficient sophistication) to a new situation. It can solve problems. It can direct conduct (as in oil refineries and space explorations; all astronauts are guided in the first, critical miles of take-off by computers that observe and correct the trajectory of the space vehicle). It can even answer certain questions from standard intelligence tests. In these matters, the observed behavior of computers and men differ not in kind, but in degree.” (Donald G. Fink, *Computers and the Human Mind*)

*22. (Letter to the Editor:) “Reading your article on Sacramento’s deplorable welfare situation brings to mind my college history professor’s lecture on the ‘Rise and Fall of the Roman Empire.’ He allowed as how the Roman Empire became so liberal and supportive of the poor that finally everyone wanted to be “poor” and live off the government. The government gave the people no incentive to support themselves. The liberal lawmakers wanted to control the people’s minds and lives. They did this by doling out most of the taxpayers’ money to welfare. The empire’s military might, which had brought about its rise, was now sorely neglected, which in turn brought about its fall.... could it be that Sacramento and the nation are experiencing the very same thing that happened in Rome several centuries ago?...I can only conclude that our legislators, like those of the Roman Empire, want to control our lives and our votes to the extent they are jeopardizing the freedom of this nation by spending our taxpayers’ money too freely for welfare and too little for defense.”

*23. “An electron is no more (and no less) hypothetical than a star. Nowadays we count electrons one by one in a Geiger counter, as we count the stars one by one on a photographic plate. In what sense can an electron be called more unobservable than a star: I am not sure whether I ought to say that I have seen an electron; but I have just the same doubt whether I have seen a star. If I have seen one, I have seen the other. I have seen a small disc of light surrounded by diffraction rings which has not the least resemblance to what a star is supposed to be; but the name ‘star’ is given to the object in the physical world which some hundreds of years ago started a chain of causation which has resulted in this particular light-pattern. Similarly in a Wilson expansion chamber I have seen a trail not in the least resembling what an electron is supposed to be; but the name ‘electron’ is given to the object in the physical world which has caused this trail to appear. How can it possibly be maintained that a hypothesis is introduced in one case and not in the other?” (Sir Arthur Eddington, *New Pathways in Science*)

*24. The writer Alice Walker was awarded a statuette by the California state legislature for being a “state treasure.” The legislature wanted to honor Walker for her novels and short stories, and probably atone for the fact that one of her short stories was removed from a state reading test after objections to the story were made by Christian conservatives. The Christian conservatives objected to “Roselily” being required reading for a state aptitude test on the grounds that it promoted unwed motherhood (the story presents the thoughts of a pregnant young woman just before her marriage).

Walker objected to the statuette of a nude woman’s torso. She called it an embodiment of society’s acceptance of the mutilation of women. David Link, a Los Angeles attorney, wrote an article arguing that Walker’s reaction to a statuette was very similar, if not identical, to the reaction to her story by the Christian conservatives. Link also argued that since the Christian conservative reaction was an “aggressive act of interpretative violence, worthy of the strongest condemnation,” so too was Walker’s reaction to the statuette designed by “one of the most gifted sculptors of our time, Robert Graham.” [in “Alice Walker isn’t that easy to honor,” *L.A. Times* reprinted in *The Sacramento Bee*, April 30 1994, p. B7.]

25. “Anybody who wants to repeat an experiment in modern subatomic physics has to undergo many years of training. Only then will he or she be able to ask nature a specific question through the experiment and to understand the answer. Similarly, a deep mystical experience requires, generally, many years of training under an experienced master and, as in the scientific

training, the dedicated time does not guarantee success. If the student is successful, however, he or she will be able to 'repeat the experiment.'

"A mystical experience, therefore, is not any more unique than a modern experiment in physics. On the other hand, it is not less sophisticated either, although its sophistication is of a very different kind. The complexity and efficiency of the physicist's technical apparatus is matched, if not surpassed, by that of the mystic's consciousness--both physical and spiritual--in deep meditation. The scientists and the mystics, then, have developed highly sophisticated methods of observing nature which are inaccessible to the layperson. A page from a journal of modern experimental physics will be as mysterious to the uninitiated as a Tibetan mandala. Both are records of inquiries into the nature of the universe" (Fritz Capra, *The Tao of Physics*).

Exercise 7-9

In the 19th century, many temperance societies were formed in the U.S., Great Britain, and Europe in response to increasing alcoholism (e.g., the Woman's Christian Temperance Union). Temperance workers such as Susan B. Anthony (1820-1906), Lucretia Coffin Mott (1793-1880), and Carry Nation (1846-1911) organized efforts to induce people to abstain from alcoholic beverages. The temperance movements urged prohibition, and the Prohibition Party made it a national issue in 1869. The temperance movement gained impetus during World War I, when conservation policies limited liquor output. In the U.S., temperance advocates worked for and, in 1919, secured federal prohibition through the 18th amendment to the U.S. constitution. But enforcement, through the Volsted Act, which prohibited the sale, manufacture, and transportation of all alcoholic beverages, failed to abolish bootlegging and the widespread lawbreaking associated with it. The illegal production and distribution of liquor during prohibition became a large industry dominated by gangsters. Its association with graft and violence was a major factor in Prohibition's repeal in 1933.

An analogy has been drawn between this era and the current drug situation, Drug sale and production is dominated by criminals and associated with violence and corruption said to be analogous to what happened during prohibition. Thus, it has been argued that we should repeal all laws which absolutely forbid the sale, manufacture and transportation of drugs. Just as the repeal of Prohibition ended the gangster era of Al Capone, so too would the repeal of drug laws end the street crime, violence and corruption associated with illegal drugs.

Is this a good analogical argument?

Further Reading – Chapter Seven

Giere, Ronald (2005). *Understanding Scientific Reasoning*. 5th ed. Wadsworth.

Langley, Russell (1970). *Practical Statistics Simply Explained*, rev. ed. Dover Publications, Inc.

Postman, Neil (1992). *Technopoly*. Alfred A. Knopf.

Werkmeister, W. H. (1948). *An Introduction to Critical Thinking - A Beginner's Text in Logic*. Johnsen Publishing Company.

Notes – Chapter Seven

¹ "Homosexuals and the 10% Fallacy," J. Gordon Muir, *Wall Street Journal*, March 31, 1993.

² I would recommend that every college student take a course in statistics. Statistical methods are used in most scientific and sociological fields. They are also the basis for the incessant polling and surveys done to discover and, sometimes, to sway public opinion. The material in this chapter is being presented in a simplified way in order to provide a foundation to the beginning college student for further study. No statistical formulae will be presented and all mathematical concepts will be presented in a simplified way. For the student who has not taken a course in statistics, I recommend Ronald Giere's *Understanding Scientific*

Reasoning, 2nd ed.(New York, Holt Rinehart, Winston: 1984), chapters 10 and 11.

³ For most TV, newspaper and magazine polls, if the samples were selected in such a way as to get a representative cross section of the target population, the margins of sampling error for various sample sizes would be approximately as follows: about plus or minus 4% for a sample of 500; about plus or minus 3 % for a sample of 1,000; about plus or minus 2% for a sample of 1,500.

⁴ “Making Sense of the Polls,” Warren J. Mitofsky in *Brill’s Content*, May 1999.

⁵ “The media: How opinion polls start to replace facts,” Bill Kovach, *Sacramento Bee*, September 23, 1989, p. B9. Kovach is a former editor of the *Atlanta Journal-Constitution*, and is curator of the Nieman Foundation, a Harvard program for journalists.

⁶ “Bush’s cautious alter ego,” *U.S. News & World Report*, March 26, 1990, p. 26.

⁷ *ibid.* Richard Wirthlin, an adviser to Ronald Reagan, was nicknamed “Numbers” because his usual support for policy matters came from polls which showed that the majority agreed with Reagan on most policy issues. “Public opinion was all he talked about, largely ignoring political strategy.”

⁸ “The media: How opinion polls start to replace facts,” Bill Kovach, *Sacramento Bee*, September 23, 1989, p. B9.

⁹ *ibid.*

¹⁰ Cable News Network (CNN) announced on July 23, 1990, that during its evening news broadcasts it would poll viewers to determine which stories they wanted to see. While the anchors read the news, viewers are shown a menu of topics. Viewers can call a ‘900’ number to vote for the story they’re most interested in seeing. The top two vote-getters are the stories that viewers will see that evening. Which would you rather see: a story about rich people defaulting on loans or a story about housewives who moonlight as prostitutes? Which would be your choice: a story on problems with nuclear waste disposal or a story on priests with mistresses? This feature is sure to be just as popular as CNN’s nightly polling of viewers for the opinions on such things as ‘Do you think the stock market is a good way to invest your money?’ Why don’t we get an investigative report on investing in the stock market? Are we really enriched by hearing the truck driver from Nantucket who thinks the stock market is just like gambling? Do we really benefit by the call from the Detroit stockbroker who assures us that his clients are doing real well by investing in stocks?

Fortunately, CNN realized the error of their ways and has discontinued both of these practices.

¹¹ “Call in polls: Pseudo-Science debases journalism,” by Richard Morin, *Sacramento Bee*, Feb. 12, 1992, p. B11. Morin is the director of polling for the *Washington Post*.

¹² *ibid.*

¹³ “Punishment Outside Prisons,” *Newsweek*, June 9, 1986, p. 82.

¹⁴ “The Story Behind the Stat,” Robert Schmidt, in *Brill’s Content*, May 1999, p. 38.

¹⁵ “Shouting ‘Fire!’,” Alan M. Dershowitz, *Atlantic Monthly*, (January 1989), pp. 72-74, reprinted in *The Art of Reasoning - Readings for Logical Analysis*, Stephen R. C. Hicks and David Kelley (New York: W.W. Norton & Co.: 1994), pp. 258-264.

¹⁶ A few years later (2003) in *Lawrence v. Texas*, the United States Supreme Court overturned *Bowers* in a 6-3 decision. Justice Anthony Kennedy, writing for the majority, said that in *Bowers* the Court had viewed the liberty at stake too narrowly. He wrote that intimate consensual sexual conduct is part of the liberty protected by substantive due process under the Fourteenth Amendment. Homosexuals, according to Kennedy, have “the full right to engage in private conduct without government intervention.”

¹⁷ See *Reasoning and Writing*, Kathleen Dean Moore (New York: MacMillan, 1993), pp. 166-167.

¹⁸ “The Delusion of Design and Purpose,” Clarence Darrow, from *The Story of My Life* (New York: Scribner and Sons, 1932), cited in *Philosophy and Contemporary Issues*, 6th ed., edited by John R. Burr and Milton Goldinger (New York: MacMillan Publishing Co., 1992), pp. 123-130.

¹⁹ *ibid.* p. 124.