Using Judgment Wisely

by
Paul Anglin*
University of Guelph
March 2011

Executive Summary

Judgment is an important business skill, especially if no formal source of information is available. But this skill is often abused for several reasons displayed by business students. One type of abuse is based on over-confidence, which should be an especially significant reason to students who have not finished learning. A second type of abuse is based on a lack of awareness: to use evidence without understanding its meaning or context. A third type of abuse is based on ignorance: to use judgment as though nothing else were available. Recognizing that ignorance is another way to describe unexplained variation in data leads to deeper insights about business planning and about research methods.

In my opinion, good judgment means knowing how to combine or weigh *different* bits of information. What many people do not realize is that weighing different bits of evidence uses a dimension of understanding that is often overlooked, known variously as the variance of a random variable, the degree of confidence or precision. I explore this dimension for both numerical and non-numerical types of evidence. Good judgment is often embodied in a good process because, as contrast, post-mortems on many famous failures often discover that well-intentioned but short-sighted "short-cuts" were used and that contradictory evidence was available but ignored.

These notes end with three appendices which contain tests to determine your awareness of your own sense of judgment.

---

* This is a work in progress designed to stimulate comments. Though they are blameless for any faults, I would like to thank Bram Cadsby, Sirui Guo and Paulette Padanyi for their help. If you have questions or comments, please direct them to me at panglin@uoguelph.ca or to College of Management and Economics, University of Guelph, Guelph ON Canada N1G 2W1

Sometimes, the CEO is forced to bet the company on building an entirely new product. The CEO of Boeing did so with the 747 (1960s), the 777 (early 1990s) and with the 787 (mid 2000s) and as he did by choosing to *not* compete with Airbus for the A380 (late 2000s). At that time, he does not know whether the necessary technology can be developed or whether the customers would continue to buy the plane for the next 20 or 30 years. The CEO of the Bank of America did it when buying Morgan Stanley over a turbulent pressure-packed weekend without knowing how much risk was involved with the toxic assets owned by this could-be-insolvent company. It sounds gutsy to bet a company on a hunch or a personal judgment. That behaviour seems to be what famous CEOs do and business students aspire to be business leaders.

At lower levels of the corporate hierarchy, decision makers depend on their ability to apply professional judgment since the data may be too fuzzy or too disorganized. It can seem that the requirements of any academically-respectable research method are, at best, too demanding or, at worst, can lead to missing opportunities within the deadlines demanded in business. Books like Malcolm Gladwell's *Blink* tell you that all subsequent activities disguise the impact of a first impression.

That perspective is only partially true since not all hunches are correct. The essential problem with using quick judgment to replace formal analysis is that, if your judgment is correct then it seems as though your judgment was good but, at the time that you invoke your judgment, the fact that you think it is correct does not make it correct. You should remember that that gutsy CEO probably got the job because he or she had repeatedly demonstrated the wisdom to know when it is better to wait and think about the problem some more.

Most people prefer to take an *informed* bet and this note argues that two characteristics are demonstrated by the application of good judgment:

- deciding how much weight to attach to different bits of information, and
- a consistent process (i.e. one which is independent of the final answer).

This note argues that both characteristics apply to the use of numerical and non-numerical data and, especially for students used to copying data from one place into a spreadsheet formula or trying to make sense of the mysteries of that formula, more attention needs to be paid to the quality of the information source. This perspective may help you to better understand that a research process adds value beyond the simple collection of facts.

The first characteristic may be easier to understand. Implicitly and especially in a situation with conflicting evidence or conflicting bits of data, the use of judgment is evident in deciding which bits are important and which bits of evidence are not important enough. What may be less well-understood is that a similar trade off also applies when all of the evidence is pointing in the same direction, but the magnitude is uncertain, since magnitude affects the scale of operations and other dimensions of a business strategy. Or, since some data sources are more trustworthy than other sources, the same evidence from different sources can have different implications.

The second characteristic of good judgment may be less obvious but is connected to several different questions. To some people, the depth of factual knowledge does not matter very much since, it seems, that any fact can be uncovered using Google, Wikipedia or, even

more audacious, Wolfram's Alpha. To many younger people who have used these websites to complete high school essays, this description of the world is self-evident. Unfortunately, when the questions become more complicated and more individualized, the answers cannot be found online. Business students often fail to understand that if they could be found easily then they would not be a basis for a competitive advantage. Most importantly, using a search engine is only useful if you know when to ask a question and what question to ask.[1]

Good judgment is not a kind of magic key that will open to lock to vast profits as in "I know something that you do not!". Any magic in the key would depend on complete confidence in the data, even though most people know that data are imperfect. One of the lessons from Tichy and Bennis (2007), repeated by other analysts, is that judgment is a process and not an outcome. Research and decision-making would be much easier if you already know the answer but, if you already know the answer, then any analysis would be a waste of time. Until you know the answer, it is hard to know which bit of information might make a critical difference to your decision: more than knowing (old) facts, having an effective process to find (new) facts matters. Most importantly, a careful process should not allow suspicions of the conclusion to contaminate or alter the analysis.

In market analysis classes, I am asked one question repeatedly and these notes are an attempt to give a more precise answer. The question is stated as something like "My judgment is my opinion. How can *you* say that *my* opinion is wrong?" My response is usually something like "I *can* question, or disagree with, the basis of the opinion." but, since students are usually not sure enough of the basis of their opinion, the student's question is appropriate but not very productive. Or, the student may answer my question with an ad hoc basis of their opinion which is not constructive (and may be inaccurate). These notes try to make this discussion more productive and constructive by identifying one part of the opinion-making process that the naïve question overlooks. And, since too many people are over-confident in their judgment, these notes try to sensitize you to why people mislead themselves by placing too much confidence (i.e. too much weight) on certain types of evidence.

These notes develop the argument in several stages. The first stage shows how the idea of weighting applies to numerical data. The arguments used here may be familiar to people who have taken simple statistics classes. This idea is applied in different sections to different issues, including the issues of variability, confidence (and confidence intervals), Type I *and* Type II errors and the importance of process independent of the answer. The second stage extends the discussion to non-numerical types of data and emphasizes the idea that all analysis needs to be concerned with bias and precision. Numerical methods have the advantage that such concerns can be summarized more easily while non-numerical evidence must deal with the problems of "post-diction" and the unstable meaning of words. The second last section discusses some of the sources of mistakes in judgment. Implicitly, this section argues that mistakes in judgment can be traced to a problem in weighting or a problem in the process. After some concluding comments, these notes offer three appendices which you can use to examine how well you understand your own judgment.

[1] Carey (NYT, 2009) gave an example of a military patrol where things did not "look right". Despite the government spending billions of dollars on detection electronics and surveillance, the individual patrol needs to know when to ask "does something look out of place?" and to do so when under mental and physical stress.

Some statistical stuff

Sir Humphrey[2]: You can prove anything with statistics.
Prime Minister: Even the truth.
(soon after)
PM: Your statistics are facts and my facts are statistics.

Data, or other forms of evidence, is described statistically as a "random" variable. A random variable is commonly described by two characteristics: its mean, or expected value, and its variance. Students are used to seeing these characteristics in the context of a bell-shaped Normal Distribution but these characteristics apply more generally. In general, the mean summarizes the environment that one is trying to understand, e.g. what are sales expected to be during the next quarter ("E(S)"). Variance summarizes the effects of various kinds of errors or ignorance that prevent you from knowing the answer to this question with perfect precision: almost surely, actual sales will not equal its expectation: S≠ E(S). This statistical arithmetic should be recognizable outside of a statistics class, though used less carefully, since a report which claims to forecast what the actual sales will be (x1 or x2) is not the only possibility that *could have been* reported. These notes start with using the terms in their formal sense and show how the formal ideas provide useful intuition for a broader class of problems that may be more familiarly called "judgment".

The two characteristics of random variables imply that evidence can mislead for either of two reasons: bias and imprecision. Data about sales are biased if the mean differs from E(S). For example, if forecaster 1 *always* overestimates by *exactly* 10 percent then they are biased but perfectly precise: x1= S* 1.10. This problem is easy to remedy: decreasing the forecast by 10 percent: let X^a= x1/(1.10) implies that X^a= S always.

If you do not know the degree of bias, then using the data adds risk to a decision. Variance means that, even if forecaster 1 reports that the answer is x1= 10,000, you recognize other possible messages could have been reported: they could have said x1= 11,000 or x1= 9500. It would be nice to pick the one right answer but the logic of statistics forces to admit the possibility that neither is exactly right. Similarly, even if two forecasters produce the same prediction, i.e. x1= x2= 10,000, that fact does not mean that either is accurate. If the two forecasters agree on the same prediction then the precision of a combined prediction should increase but logic implies that *both* may be wrong. If the forecasters disagree, then it becomes even more obvious that you need to learn how best to combine the more or less wrong forecasts.

Hard numerical data can be combined using a *weighted* average. Suppose that you have sales forecasts from two different forecasters (either consultants hired as forecasters or informal forecasters like sales reps), that you want to use to make a forecast about future sales. If one forecaster says sales will be x1= 10,000 and the other forecaster says that sales will be x2= 8,500 then a weighted average would be

X= a* x1+ (1- a)* x2= a* 10,000+ (1- a)* 8,500

for some a between 0 and 100 percent (i.e. a between 0.00 and 1.00). If the first forecaster is

[2] This quote is from a British TV show called *Yes, Prime Minister* (http://www.yes-minister.com/ypmseas1a.htm, an episode concerning cigarette smoking).

more trustworthy or more precise then a should be larger and if the second forecaster is more trustworthy then 1- a should be larger.

Optimizing "a"

Adjusting reports for a known bias is easier than adjusting for differences in precision. With two unbiased reports, the primary effect of changing a is to change the precision of the weighted estimate. Therefore, the weights are optimal if they produce the *most* precise estimate of E(S). Statistical theory reveals what is needed to make the adjustment and one of the ways in which judgment is abused is because people do not always know what is needed or do not communicate that knowledge to others.

Suppose that you have two (unbiased) reports on future sales: x1= 9000 and the other is x= 10500. If you knew that sales will be S= 10000 then it would be better to let a= 1 (i.e. X= 10500). But if you knew that S= 8500 then it would be better to let a= 0 (i.e. X= 9000). The problem is that you do not know whether S= 10000 or 8500 (or any other value) until later.

An alternative is to treat the forecasts equally and let a= 1/2, but that answer asserts that all forecasts are equally useful. So, the real solution to this optimization problem starts by recognizing a difference between the forecasters other than simply their reports: i.e. the precision of the forecast or some other measure of its trustworthiness. If var(x1) represents the variance of the report from the first forecaster and var(x2) represents the variance of the report from the second forecaster, before either report is seen, then a bit of work shows that the weight which produces the most precise estimate is[3]

$a^{\#}= [1/var(x1)]/([1/var(x1)]+ [1/var(x2)])$

Notice that it is *relative* variance that matters. If forecaster 1 were known to be perfect then $a^{\#}= 1$, and this formula makes sense since it implies that all of the weight is assigned to one forecaster regardless of what the other forecaster does. If market conditions change so that it is harder for anybody to forecast accurately, e.g. both var(x1) and var(x2) double, then $a^{\#}$ would not change. If both forecasters are equally trustworthy, i.e. either equally good or equally bad, then a= 1/2. If stated in relative terms, the optimal weight satisfies

$a^{\#}/(1-a^{\#})= [1/var(x1)]/[1/var(x2)]= var(x2)/var(x1)$.

1/var(x1) is sometimes referred to as the "precision" of a random variable.

A bit of algebra shows that the precision or variance of the weighted average is[4]

$var(X)= var(a^{\#}* x1+ (1- a^{\#})* x2)= (a^{\#})^2 var(x1)+ (1- a^{\#})^2 var(x2)$

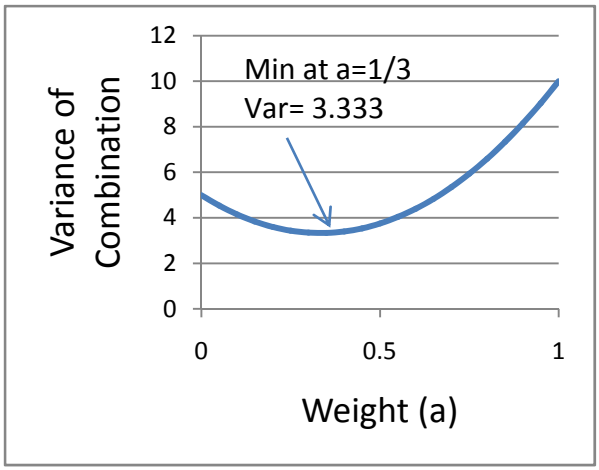
$= 1/([1/var(x1)]+ [1/var(x2)])$

In other words, combining the forecasts optimally implies that the precision of the optimally weighted average is equal to the sum of the two precisions and, therefore, must be more precise

[3] Formally, this analysis uses the Normal distribution. Using a fancier distribution might be more accurate for certain kinds of problems. Using a fancier distribution may also help because, for certain kinds of problems, both the mean and the variance vary with S. Statistics classes show how to extract as much insight as possible from the limited data available.

[4] This calculation assumes that x1 and x2 are "statistically independent", conditional on S. If that assumption is doubted, then I leave it to the reader to revise this formula to account for the covariance between x1 and x2 and I leave it to the reader to investigate when the revised formula makes a difference in terms of the business strategy.

than using either forecast alone. The figure below illustrates this fact for an example where var(x1)= 5 and var(x2)= 10 and optimizing the weights reduces the variance of X by one-third.



This calculation implies that knowing how to combine information is valuable even if it does not change the prediction on average.[5] In fact, this logic is useful because any knowledge would contaminate the process and make it harder to find the right answer consistently.

This statistical story offers a fairly accurate description of the differences between a student and an expert. Everybody has access to outside sources of information (i.e. x2) which are available for purchase. Experts have experiences which give them access to more sources and types of information (i.e. x1). This experience leads them to think that what they believe is fairly accurate (var(x1) is small) before they see any outside information. A typical student does not have this advantage and may prefer to put little or no weight on any one report, but ignoring all reports is illogical: it is impossible for both a= 0 *and* 1- a= 0 to be true. A more practical bit of advice is to recognize that var(x1) may be close to infinite for a student (i.e. 1/var(x1) may be close to 0 and $a^{\#}$ may be close to 0). Thus, it is not surprising that students tend to rely more on any outside information without regard for its precision and that this situation causes problems.

Notice that the value of a does not depend on the reports (i.e. 10,000, 11,000 or 8,500) since judgment is invoked to enhance the reliability of the combined forecast that will be used to make a decision. A change in the unknown S should change the expected predictions because the reports are supposed to reveal S. A change in S should not change the process which is summarized here by a (unless a change in S also changes var(x1) or var(x2)).

var(x1) or var(x2) can vary for many reasons, including differences in
sample size
care in data collection

[5] I can imagine a busy executive arguing that they are too busy to consider such philosophical issues of optimal weighting: "I want to pick one of these two. Only one. And then act on it". Consider three responses. First, for a student at university, their time can be spent asking and investigating the kinds of questions needed as preparation to succeed as that busy executive. Second, and unless you already know the answer, there is a saying worth remembering: "act in haste, repent at leisure." Third, the fact that a busy or successful executive says something does not mean that others are convinced by what they say: that executive knows that he or she needs to convince others who have seen the other source(s) of information.

sample selection (e.g. internet poll vs. census, or narrow vs. national coverage)
time period (both time span and age).

Statistics classes and research methods classes spend a lot of time discussing the impact of such issues and ways to correct for such concerns. Differences in the business environment would also change the amount of intrinsic variability and affect data quality: e.g.

new businesses tend to face more risk (compared to mature industries) or
local markets tend to be more familiar than markets far away.

It may be obvious that a business strategy depends on judgment which, I argue, depends on the variability of the data used for evidence. The choice of strategy, especially when focusing on a segment of consumers, may have an indirect effect of lowering the quality of the data available to fine-tune the strategy.

There is a second way to see the significance of a. In many cases, and even for high quality data sources, the published data are revised at a later date. At one level, the fact that a report can be revised without changing reality should demonstrate a meaning of precision. At a second level, the example of revising the data reveals the impact of judgment: if $x2^{\#}$ had been reported instead of x2 then the person using the report from forecaster 1 should have been comfortable with a judgment that

$$X^{\#} = a * x1 + (1 - a) * x2^{\#}$$

instead of

$$X = a * x1 + (1 - a) * x2.$$

Therefore, the *difference* in judgment caused by the revision would vary with a in an obvious way:

$$X^{\#} - X = (1 - a) * (x2^{\#} - x2).$$

This kind of sensitivity analysis also enables people to investigate the basis of a conclusion without directly questioning the conclusion of an analysis. For a student who uses a close to 0, the answer for X changes with the realized value of x2 in almost a one-for-one basis.[6]

How do these ideas apply to non-numerical bits of information?

> "Everyone is entitled to his own opinion," the great Sen. Daniel Patrick Moynihan of New York was fond of saying. "He is not entitled to his own facts."[7]
>
> Q (in a disdainful English accent): "Good to see you, Mr. Bond. Things have been awfully dull round here.

[6] A particular application of this logic may be seen in classroom discussions of a case study. Well-written cases can give a reader a sense of the problem in a few pages and students who see many cases begin to see patterns or common features in how to approach a problem. But, it is also possible for a student to use the writing style to decode the focused material and offer an answer with limited personal insight. This fact represents a limitation of the case study method for a weak student since the case study is intended to be the sole source of available information. Therefore, any change in the information in the case would change what a decoding student would write in a one-for-one way. Only during discussion and debriefing is there an opportunity to fairly and accurately introduce other sources of information.
It would be interesting to find a way for a student to add judgment explicitly as an independent information source, if a way could be found to make this source independent, including being independent of information which becomes available after the events written in the case.

[7] The first quote can be found at http://www.popularmechanics.com/technology/military/news/1230517 and many other places. This link may also serve to illustrate the consequences of a research process which is well-intentioned but careless.
The second quote is from http://www.movietranscriptions.com/60377_Never_Say_Never_Again.html (*Never Say Never Again*, 1983).

> Bureaucrats running the place, things done by the book. Can't make a decision unless the computer gives you the go-ahead. Now you're on this, I hope we're going to have some gratuitous sex and violence." James Bond: "I certainly hope so too."

Many types of useful information are hard to translate into numerical data without losing something in the translation. You may not even be aware of this continuing flow of relevant information also known as rumours, "soft information" or a vague memory that you heard somebody say something important to somebody else. For this reason and for other reasons, some people are uncomfortable using numerical data to justify a decision and may use the concept of "judgment" to introduce other types of information. In principle, decisions should be based on as much information as possible and, where the previous section used the example of numerical data to motivate the idea of judgment, this section suggests that the same concerns with bias and precision need to be considered. In addition, the use of non-numerical evidence creates a special issue: independent of any uncertainty about the state of the world, the meaning of words may not be stable.

Obviously, it is not sensible to state or use a weight of a= 0.333 or a= 0.546 when combining one or two bits of non-numerical information. But, the issues raised in the previous section can be translated to convey a comparable conclusion. First, even if not measured in terms of statistical variance, the precision of an information source needs to be accounted for: for a given set of market conditions or a given value of future sales, what are the possible reports which could be seen (before the report is seen)? Which reports are more likely to be seen and what kinds of reports would be regarded as extreme or unlikely? Information should be seen as more imprecise if the reports are more diverse or more variable. Second, even if a weight is not measured directly, it can be inferred by noting the effect on X of a change in the information, as noted in the last paragraph of the previous section.

Third, as noted in the box above, it is not necessary to measure the precision of every data source separately: the optimal weights depend on the *relative* precision of the different information sources. Even if not precision is measured in any formal sense, it is possible to describe the kinds of situations where the connection between the reported information and the underlying truth varies in a one-to-one basis and the kinds of situations where the connection would be weak. Any one truth may be reported (with error) in more than one way and any bit of reported information may be the result of more than one possible truth. In a mature industry with little change, there would be few confounding influences and any information that is available would be more precise than if the same information had been reported about an emerging market by somebody who is known to lie occasionally. So, implicitly if not explicitly, the idea of weighing different bits of evidence is relevant but challenging.

Part of the challenge may lie in a difference between numbers and words or, as some people call it, between "hard data" and "soft data". Numbers are numerical and, except for some debates about the differences between Arabic and Roman numerals, their meanings are not debated much.[8] In contrast, the meanings of words are often debated. Office games like Bullshit

[8] Government statistics are revised sometimes and one can debate whether the right numbers are collected, but there is no debating the difference between sales of "35.4 units per week" and "34.5 units per week": if one is true then there is no uncertainty about the other being false. These notes discuss the significance of a difference, either in terms of statistical significance or in terms of having a significant effect on an operational decision.

Bingo which are funny precisely because many important-sounding words have been abused to the point of being *meaning*less; and the audience knows it. People use code words and weasel words that don't always mean what you think they mean:

"affordable"
"reasonable" and "fair share"
"enough"

Politicians can exploit the unstable meanings of words and political activists on different sides of a debate often use different words to assign different meanings to the same thing:

"pro-life" vs. "pro-choice"
"terrorist" vs. "freedom-fighter" or
"level playing field".

While not all information is displayed numerically, abuse of non-numerical information should not be tolerated.[9] Appendix 2 offers a test which you may or may not convince you and a perspective on how to avoid the problem.[10]

People use words to confuse or are confused by words in three different ways: logic, margins of adjustment and probability. First, and independent of whether enough information is available, many people cannot recognize whether a series of statements is sufficient to establish the stated conclusion. To give an obvious example,

"Terrorists are Muslims.
Iraqis are Muslims.
Therefore, Iraqis are terrorists."

Although this type of fallacy is routinely ripped to shreds in any formal analysis of logic, the final statement may seem obviously true to the audience when dressed up in more colourful or emotive language by a skilled speaker. This logic is not uncommon:

Advertisements are deceptive.
Marketers make advertisements.
Therefore, marketers are deceptive.

or

Marketers are deceptive (which is bad).
Marketing instructors teach marketing.
Therefore, hiring marketing instructors is bad for society.

Many people in marketing would object to the conclusion in each case even if they agree that the first two statements might be correct. It would be more helpful to clarify the statements with modifiers, like "some" or "many" or "a few" as an obvious entry point for the use of more information.

---

[9] If the meaning changes based on knowledge of the outcome (i.e. which is possible only when it is too late to act) then the words convey less operationally-relevant information than expected. The essential confusion is between "post-diction" and "prediction". Post-diction gives the appearance of predictive wisdom because the outcome seems to match the words used but the match is made easier by using vague words. Centuries of fascination with Nostradamus and with modern psychics demonstrate its seductive power. To see the tricks played by "psychics" to create confidence where none is justified, go to http://skeptico.blogs.com/skeptico/2007/11/john-edward-jam.html

[10] Another example of verbal vagueness may be familiar to job applicants. People are often advised to offer specific accomplishments: instead of saying "I did a good job" and "I am a good communicator", it is better to say "I increased sales by 10 percent" or "I won an award for public speaking". Recruiters at large companies are aware of how the vague language has been abused by previous applicants.
See also the small book by Frankfurt (2005) entitled *On Bullshit*.

A course in formal and informal logic would show why and when the third statement is not properly connected to the first two.[11] Many business people will rightly claim that they are too busy to take such a course, especially if it were taught by a philosophy instructor instead of a business instructor who is expected to be more practically-oriented. This claim implies that business people are willing to knowingly make incorrect conclusions. For this reason, business students should use the luxury offered by their time at university to learn how to avoid the problems of their elders. Or, to modify a familiar platitude: Develop your judgment now because, later, it will be hard to remember that your job is to clear the swamp because you are up to your ass in alligators.

The careless use of words can create problems in a second way when they describe margins inaccurately. For example, people may think there are only two types of answers (true or false), or a consumer is either willing to pay at this "price point" or not or there are only two types of precision (0 or perfect).[12] Business gurus encourage this belief by using various tools, such as 2x 2 matrices where all companies can be divided into one of the four quadrants. A binary distinction is false: there are in-between cases. Simple categorization may clarify thinking in the initial phases when identifying the real question or when preparing to make a judgment, but the categories also hide distinctions which may be decision-relevant. So, some careful translation of the words is needed before judgment can be applied.

---

[11] For a list of logical fallacies, please see Wikipedia. The fallacy illustrated in the quote above is known as the "undistributed middle": http://en.wikipedia.org/wiki/Fallacy_of_the_undistributed_middle
For an entertaining ripping-apart of bad logic, see http://www.thedailyshow.com/watch/thu-march-18-2010/conservative-libertarian or, in Canada, http://watch.thecomedynetwork.ca/the-daily-show-with-jon-stewart/headlines/the-daily-show-with-jon-stewart---march-2010/clip280324#clip280324; if you think of yourself on the political right of Glenn Beck, note the appalling style of logic and, if on the political left, pay extra attention to the logic since you probably already agree with anything Stewart says regardless of whether it is logical or true. To see where the logic of "associated with" and "linked to" (common reported in media analyses) can go, please see http://www.thekevinbacongame.com/ If you think about it carefully, paying close attention to the variety of different kinds of "links", it would be surprising if *you* were not "closely linked to" somebody evil.

[12] Consider two paragraphs:

> A: "In the worst case, deflation becomes its own cause. People become afraid their incomes might fall in the future. Or they see their savings being ravaged by the stock market collapse. So they stop spending and instead hoard their money. As demand for goods and services drops, companies' profits plummet, leading to layoffs, reduced working hours, and yet more declines in stock prices. The fear of lost income becomes a self-fulfilling prophecy, and people cut their spending further. Once the downward spiral starts, it's maddeningly hard to stop. People expect prices will keep falling, so they decide to put off their spending, because they think things will be cheaper in the future."
>
> B: "Inflation becomes its own cause. People hope that their incomes might rise in the future. Or they see their savings boosted by the stock market expansion. So they spend more. As demand rises, companies' profits increase, more jobs, and more increases in stock prices. The hope becomes a self-fulfilling prophecy, and people spend more. Once the upward spiral starts, it's *maddeningly hard to stop*."

The first is taken from a well-known "great thinker" (in *Globe and Mail*, Nov. 24, 2008. p. A17 http://www.theglobeandmail.com/servlet/story/RTGAM.20081121.wcoecon24/BNStory/specialComment/). The second paragraph is the same as the first except that every example of "down" is reversed into "up" and vice versa. Version B would have been "obviously" true to most readers if it were published 18 months prior to version A. We now know that something was missing from version B and, despite its seductive appeal, something similar must be missing from version A.
Neither paragraph should be convincing once you start to think about what information would be needed to make an argument convincing. Attempting to quantify something, even if the best attempt is imperfect, is often seen as useful because it reveals missing bits of the puzzle.

Another important class of margins creates a puzzle that is too complex to solve here. Words are very good at expressing certain ideas and not others. Consider the following statement: "__ people will buy our product." That blank can be filled in with words such as "no", "a few", "some",[13] "half", "average" "many", "most", "nearly all" and "all" but that blank needs to be interpreted if you are making the kind of sales forecast that will be used to determine whether you are to be paid a bonus. And, if somebody buys, does each consumer buy one and only one unit or do some buy one unit per year while other consumers buy 2 units per week. More generally, words are good at ranking items between "more" and "less" but numbers are better if you need a cardinal scale.[14]

Degree of Confidence

> "That's right" shouted Vroomfondel, "we demand rigidly defined areas of doubt and uncertainty!"
> Douglas Adams, *Hitchhiker's Guide to the Galaxy*

Another way to distinguish the different dimensions of judgment is to introduce an idea that is familiar from statistics: a confidence interval. Where weighting shows the importance of some evidence or bit of data *after* it becomes known, a confidence interval shows the importance of one outcome by identifying a range of outcomes that could have been seen. For example, when asked for an estimate concerning a question, such as "How much did Canadians spend on watermelons last year?", people may answer with a specific number "one million dollars" but the answer usually starts with something like "about …" to show that the number is not known precisely. If asked for an interval of likely answers, a careful expert would say something like "between $0.9m and $1.1m" while a novice might answer the same interval question with an answer like "between $0.5m and $2m". If a news report noted that sales were $1.2m then this outcome would be surprising to the expert but is not unexpected to an ordinary person.

|----------|
A 50% Confidence Interval for Person A

|--------------------------------|
A (Wider) 50% Confidence Interval for Person B

|--|
A 95% Confidence Interval for Person A

A confidence interval is range which passes a predefined reasonableness criterion: scientists often use a criterion such as 95 percent, media reports on an opinion survey may use a phrase like "19 times out of 20" (where 19/20= 0.95), the banking sector uses this kind of idea in the form of a "99% (daily) Value at Risk", but others may use a different standard such as 50 percent. Since a person's judgment is not intended to be answered as either "Yes" or "No", the

[13] For example, "some people say …" but, to show the weakness of this type of argument, Wikipedia notes "Other people say the opposite".

[14] Operational decisions often invoke numbers: the price is not "high" but a precise number and a production target is not "a lot" but a specific number of units per hour.
While unsubstantiated rumours are too vague to be quantified onto a scale or more or less, it may be possible to translate the effect into something measureable and operational.

use of an interval allows the expression of judgment to vary according to the question and according to the degree of confidence you want to place in an answer: e.g. a 50 percent confidence interval to a question with an "obvious" answer would be smaller for two reasons than the 95 percent confidence interval to a question asked of the same person where the answer is as obvious.

Giving a range on what a report could say, before seeing the report, implies how surprising a report could be when it is seen. Using the numbers above, a report which says that watermelon sales were $1.5m this year would be surprising to the expert. To the novice, the same report lies inside the interval and, therefore, can be considered as not abnormal or unreasonable.

This idea of a confidence interval has many other applications and implications. First, and perhaps most important, it recognizes the limited value of summarizing an estimate by a single number or "point estimate'. I think that most students recognize this limitation and it may indicate why many people distrust forecasts. Most students recognize the problems with relying on a point estimate but, despite having taken appropriate classes, they do not take the next step to ask about a confidence interval (in any form).

This representation offers practical value to managerial decision-making. At the simplest level, checking how often the true answer lies within an interval enables you to check whether your judgment is biased *and* whether it is as precise as you think. Judgment is biased if there is a pattern of being too high or too low on average. The managerial implications of using a consistently high or consistently low estimate should be obvious.[15]

A person's judgment is *excessively imprecise* if the true answer is always inside the confidence interval too often. Choosing an interval (with less than 100 percent confidence) has measureable consequences since the true answer should be surprising sometimes (i.e. lie outside of the chosen interval). If the true answer always lies inside the interval, then it displays a pattern of thinking that an event is reasonably likely (where "reasonable" is defined by the stated level of confidence) even though the pattern shows it is not that reasonable. In other words and even if your judgment is right on average, excessive imprecision implies that the chosen best case scenario (i.e. the upper bound on the interval) is too optimistic and the chosen worst case scenario (i.e. the lower bound) is too pessimistic. The costs of this pattern of poor judgment are evident in the development of contingency plans which are never used.

Displaying an *excessively precise* judgment is not a good idea either. If the true answer lies outside the confidence interval more often than the stated degree of confidence, then you will be surprised more often than you expect and the business strategy is unlikely to be matched to reality.

[15] These statements are very broad and a technical caveat may be obvious to you: though asked a range of questions, you may be an expert on certain types of questions and a novice on other types of questions. So, on questions for which you are an expert, your confidence interval should be smaller than if you were a novice. The lesson to learn from the idea of a confidence interval is that mistakes are a problem when there is a pattern: anyone can be lucky or unlucky once.

The second advantage of thinking in terms of confidence intervals is that it is a way to put a price on unrealistic claims. For example, Bazerman and Gillespie (1999) argued that some negotiators use a tactic where they claim to worry about a wide range of possible outcomes. They suggested that you need to distinguish between sincere beliefs and claims made for the purposes of extracting some bargaining surplus. They proposed that you should make them pay a price for insincere claims: for example, make the terms of the contract depend on the outcome which they claim to worry about. Basing negotiations, or any analysis, on only a point estimate or only on an average outcome ensures that the basis for the negotiations is not matched to reality of the situation, limits your awareness of what the other person knows, and limits how that knowledge could help you and them.

Even within an organization of a team, "putting a price on unrealistic claims" adds structure to the idea of "best case" and "worst case" scenarios when talking about contingency plans. The essential problem with the phrase "worst case" is that the "worst" that people can think of is usually not the worst that could possibly be:[16] in the real worst possible case, the outcome is terrible and there is nothing you can do about it. In practice, what is needed is a "reasonable" worst case scenario. Where the word "reasonable" suffers from the vagueness criticism noted above, the use of a confidence interval offers a precise sense of the degree of "worst" and this precision has operational implications.[17]

Confidence intervals offer a way to resolve issues of self-delusion and cynicism. As noted in the next section, research has shown that, if asked to give an interval, many people tend to give intervals which are smaller than they should be. Opposing this research is an attitude, which seems prevalent amongst cynical business students who are tired of the restrictions of school-based learning, to assert that "anything could happen": if asked to give a confidence interval about future sales where "anything could happen" then their 99.9% confidence interval could stretch from nearly 0 to nearly infinity (say $10b).[18] A way to start to resolve both types of inaccuracy is to think about "putting a price on unrealistic claims". For example, if the business student's assertion is sincere then they should be willing to "put their money where their mouth is" in a more challenging test: instead of asking for a 99.9% confidence interval, ask for something else like a 50% confidence interval or a 70% confidence interval. This revised question does not challenge the assertion directly but the narrower question offers more insight into which possibilities are more likely. Appendix 1 offers a test which you can use to judge whether your judgment on a series of questions related to business is biased, excessively precise or excessively imprecise.

---

[16] http://en.wikipedia.org/wiki/The_Worst-Case_Scenario_Survival_Handbook

[17] The most important operational implication is that you should *use* numerical forecasts but not *rely* on them. The only way to "use but not rely on" a forecast is to understand its precision. For example, every forecast that reports a point estimate is almost guaranteed to be wrong: if a forecaster predicts that growth will be "3.5 percent" then the actual outcome will be something other than 3.500000 percent. A confidence interval around that estimate would show which outcomes are more likely and offer a more useful basis for developing contingency plans. Alternatively, you can investigate the context using "What If" analysis and Sensitivity Analysis.

[18] A more advanced version of this unconstructive assertion is "anything could happen, independent of what we decide". Though this assertion notes the relevance of conditions in any statement of judgment, using judgment in this case is irrelevant since the statement asserts that what happens is independent of the decision. More advanced analysis could consider whether the degree of uncertainty is affected by the conditions a lot or a little.

Mistakes in Judgment

"Self deception proves itself to be more powerful than deception"[19]

"The first principle [of Science] is that you must not fool yourself- and you are the easiest person to fool. So you have to be very careful about that. After you've not fooled yourself, it's easy not to fool other scientists. You just have to be honest in a conventional way after that."

"The difficulty lies not so much in developing new ideas as in escaping from old ones."

A traditional view of judgment is that it is used at a single moment by a heroic leader who has a magic key. In reality, Tichy and Bennis (2007) argued that any magic in the key may not be seen until later and that life is not based on a single moment. It is also true that the hero does not always need to slay a dragon heroically: decisions made and judgments invoked while preparing for the supposed single moment may cause the dragon to not attack. To avoid a mistake in judgment, you need to change your point of view.

Mistakes are rarely intentional. Consequently, much of this discussion focuses on developing a good process and self-protection. Yet, mistakes are not uncommon and research (e.g. Tichy and Bennis, 2007; but see also Lovallo, Vigerie, Uhlaner, and Horn, 2007) has noted several types, which I summarize here as being evident in either bias or in precision:

confirmation bias (bias)
overconfidence (mostly precision)
planning fallacy
conflict of interest (bias)
winners curse (bias and precision)
anchoring (bias)
sunk cost fallacy (bias)

It is easy enough to remember this list and to repeat it on a test. Unfortunately, that ability is not sufficient since, at the time when you want to use judgment, you may forget the list. Thus, the repeated encouragements to be skeptical, even of oneself. Such self-awareness can be especially difficult, and important, in a competitive business environment when under pressure (see also footnote 1).

Looking for disconfirming evidence is widely advocated as a solution to bad judgment (Larrick, 2004). It can be more important to look for disconfirming evidence, even if not found, than to look for and find confirming evidence. At the grand level, this idea can be seen in the famous NASA picture of what was expected to be mostly empty space: by finding what was not expected, it changed the questions that people asked.[20] Information has non-linear effects since,

---

[19] The first quote comes from *Scientific American* http://www.sciam.com/article.cfm?articleId=E7327616-E7F2-99DF-38F214BFD77FE010&chanId=sa013&modsrc=most_popular (May 2007) . See also http://www.skepdic.com/selfdeception.html. The second quote comes from *Surely you're joking, Mr. Feynman: Adventures of a curious character*, R. Feynman (Nobel Prize winning physicist) as told to R. Leighton, Vintage Press, 1992 (p. 343). The third quote comes from the famous economist John Maynard Keynes (http://thinkexist.com/quotes/john_maynard_keynes/)

[20] Another common source of mistaken judgment is prejudice: you are so confident in your judgment that you pre-judge the truth. By not questioning "conventional wisdom", you may overlook an opportunity to gain an advantage over competitors who remain constrained by conventional wisdom. The fact that the NASA picture was taken at

while you may not expect to find anything, if you find anything then what you find may be important at a fundamental level.

At a practical level, post-mortems on famous failures often show how negative evidence was ignored. In an uncertain environment, an outcome depends on a mixture of luck and skill. Thus, the outcome is an imperfect indicator of the quality of judgment. This fact is important because it is easy enough to fall into a trap of thinking you have good judgment by using a selective memory: if a good event happens then it is because of skill and if a bad event happens then it is because of bad luck. The value of doing a post-mortem for a "failed" project is relatively obvious. Doing a post-mortem on a project with a good outcome has many benefits:

identifies whether the good outcome could have been better,
establish a baseline of what is "normal",
communication within the organization and information sharing and
reduces any stigma associated with doing a post-mortem only on "failures".

During the post-mortem after a disaster, investigators often find that good processes were contaminated by people who effectively assumed that the thing which caused the disaster could not happen. "Consistency is no accident" makes sense in both meanings of the phrase.

Sometimes, a good process is not a matter of looking for more data but of asking the right question of the available data. During the post mortem after many famous failures, people go back through the data that was available at the time and see that the problem was predict*able*: a predictable surprise (HBR, Watkins and Bazerman, 2003). The failures happened because people did not understand the significance of that data or act on it. Taleb (2007) called this behaviour failing to see the *significance* of the black swan. There are well-known examples of people who did react to something surprising: penicillin was discovered after a messy lab scientist noticed dead bacteria in a Petri dish; Viagra was developed after scientists noticed a side effect of a drug developed to treat angina. To use a very old quote: "Chance favours the prepared mind" (L. Pasteur).

Some people (Alloy and Abramov, 1979; Southey and de Meza, 1996; Seligman, 2006) argued that pessimists have a useful role in an organization because they are more sensitive to disconfirming evidence. (I think that it would be interesting if somebody could find evidence showing whether marketers are naturally optimistic because it would indicate using judgment to resolve a marketing question places them in greater danger.) Unlike a constant critic, a constructive pessimist wants more insightful data or adds comments such as "To convince me, you need to do the following …". This idea of constructive pessimism may be most useful when something is *half* true: as noted in the quotes at the beginning of this section, self-delusion about the whole truth can be very powerful. Knowing how to disprove something shows a clear distinction between proof and confirmation: "proving" that X is true also implies that "not X" is not true while "confirming" that X is true allows for the possibility that something else can also

---

great expense (using the Hubble space telescope) on somebody's discretion (rather than as the result of a specific scientific research project) over many days of an *extremely* small fraction of the sky should reinforce the lesson that investigating doubts can be fundamentally important. See http://hubblesite.org/newscenter/archive/releases/1996/01/text/ for more information.

be confirmed to be true.[21] Many hypotheses can be confirmed in a limited sense by using low quality data or a weak test or, by following the human tendency to modify a test to fit a preconceived idea.

Concluding Thoughts

> " 'I checked it very thoroughly,' said the computer, 'and that quite definitely is the answer. I think the problem, to be quite honest with you, is that you've never actually known what the question is.'" Douglas Adams, *Hitchhiker's Guide to the Galaxy*, after the computer announced that *the* Answer to "the Ultimate Question of Life, the Universe and Everything" was 42.

Using personal judgment seems like a quick way to find an answer. These notes try to suggest that this perspective is too narrow: judgment is not a magic key whose basis is unquestionable. Broadening the perspective introduces a number of issues that should be considered because they produce more reliable answers and have operational implications.

Mostly, these notes have focused on facts and more intense preparation would enable you to learn more facts to use when forming your judgment. The three Appendices offer a way to start to teach yourself by making you aware of your performance on some factual questions. If it were just a matter of knowing facts, younger people tend to not expand this background knowledge by not following the news and older people tend to follow a preferred news source which covers issues selectively. Judgment is evident when weighing different bits of factual information and this idea introduces questions about how the weights depend on the reliability or precision of the information source. Even when not using numerical data, different bits of evidence should be weighted based on measures related to variance, variability and measures of surprise. If there is no alternative to using judgment, careful thought often reveals alternatives that *should have been* explored. You do not need to *rely* on a hunch if you have the time to do some preliminary research: that is the role of a planning process. And, if you do not have the time to do the research, there are times when the available information is so imprecise that you should wait to collect more information.

Therefore, in addition to facts, process is important. The need to combine and manipulate information should explain why so many university business classes talk about research design, as opposed to research findings. These classes are useful because they describe the rigors of a good process, and help to avoid the mistakes caused by post-diction. So, a simple

---

[21] This difference between "proof" and "confirmation" opens a much longer and older debate in logic, and is especially problematic when the evidence is affected by a random variable. To give some operational implications:
a) a proof is always deemed to be temporary since evidence may be found, in the future, which proves the opposite.
b) knowing the "power" of a test indicates how much confidence can be placed in a result which says "Yes". (The "Neyman-Pearson Lemma" proves that there exists an inescapable trade off between the probability of making a "Type I error" and a "Type II error".)
c) as a consequence of a) and b), if it is possible to confirm one hypothesis, there will be similar hypotheses which which can be confirmed using the *same* test procedures. If so, do you know what those other hypotheses are and, before basing a business plan on one of them, do you how likely each is true? If more information is needed to distinguish the hypotheses, do you know what kind of information would be most powerful?
Perhaps the most recent and high profile example of this difference is in the concept of a "long tail" which was asserted to be true in a best-selling book before being critiqued http://en.wikipedia.org/wiki/Long_Tail#Criticism. The concept is rarely discussed now. Juslin et al (2000) offers a more obscure example which is both relevant to the issues discussed here and shows that scientists are not immune to making this mistake.

conclusion of these notes is that you should learn to be precise, even about what is not known, and you should learn to look for precision as a relevant aspect of the analysis.

Bibliography

Alloy, L.B., and L.Y. Abramson, 1979. "Judgment of contingency in depressed and non-depressed students: Sadder but wiser": *J of Exper. Psych: General*, 1979, 441- 485.

Bazerman, M., and J. Gillespie, 1999. "Betting in the future: The virtues of contingent contracts", *Harvard Business Review*, Sept., 155-160.

Carey, B., 2009. "In Battle, Hunches Prove to Be Valuable", *New York Times*, July 28, 2009.

Frankfurter, H., 2005. *On Bullshit* Princeton University Press, Princeton N.J.

Granger, M., and M. Henrion, 1990. *Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis*, University of Cambridge, New York. HB 615.M665

Hammond, T. R. Keeney, and H. Raiffa, 1998. "The hidden traps in decision-making", *Harvard Business Review*, Sept., 47-58

Juslin, P., A. Winman and H. Olsson, 2000. "Naive empiricism and dogmatism in confidence research: A critical examination of the hard-easy effect" *Psychological Review*, 107 (2), 384- 396.

Langer, E., "The psychology of chance", Ch. 6 in *Risk and Chance*, ed. by E. Dowie and P. Lafrere.

Larrick, R.P., 2004. "Debiasing", ch. 15 in *Handbook of Judgment and Decision Making* ed. by D. Koehler and N. Harvey, Blackwell Publ., Oxford.

Lovallo, D., P. Vigerie, R. Uhlaner, J. Horn, 2007. "Deals without delusions", *Harvard Business Review*, December, 85 (12), 92- 102.

Russo, J. E. and Schoemaker, P. J. H., 1990. "The overconfidence quiz," *Harvard Business Review*, Sept./Oct. 1990.

Soll, J., and J. Klayman, 2004. "Overconfidence in interval estimates", *Journal of Experimental Psychology: Learning, Memory and Cognition*, 30 (2), 299 –314.

de Meza, D., and C. Southey, 1996. "The borrower's curse: Optimism, finance and entrepreneurship," *Economic Journal*, 106 (435), 375-86.

Seligman, M., 2006. *Learned Optimism*, Vintage Press, New York, p. 107- 112.

Taleb, N., 2007. *Black Swan*, Random House, New York, (see also http://www.fooledbyrandomness.com/)

Tetlock, P.E., 2005. *Expert political judgment : How good is it? How can we know?*, Princeton University Press, Princeton. JA74.5 .T38 2005

Tichy, N., and W. Bennis, 2007. "Making Judgment Calls", *Harvard Business Review*, 85 (10), 84- 93.

Watkins, M., and M. Bazerman, 2003. "Predictable Surprises: The Disasters You Should Have Seen Coming", *Harvard Business Review*, March.

Weinstein, L. and J. Adam, 2008. *Guesstimation: Solving the world's problems on the back of a cocktail napkin*, Princeton University Press, Princeton, QA 276.8 W45 2008.

Appendix 1: Different Ways of Expressing Judgment

Context

You are invited to answer a set of questions about judgment and estimation. Consider the following scenario: you are about to propose a project to the Vice-President. You supervised the people who worked on the proposal but only they know all of the details. They propose a specific business strategy that seems good but the outcomes is uncertain. The Vice-President will ask you to state your confidence in the outcome because the Vice-President has a limited budget for new projects and the Vice-President wants your sincere opinion on what would be the best use of resources for the organization.

We start by agreeing that you may not know an answer precisely, and it would be better if you knew more, but it is unreasonable to expect any one person to know everything. It is reasonable to expect that you have insight, opinion or judgment concerning which outcomes are more or less likely to be true and the Vice-President is asking only for your honest opinion. The questions below investigate two ways of expressing a judgment.

The five "Q" questions suppose that the Vice President has asked you to pick intervals. Since it is not expected that you know the true answers to all questions, company policy dictates that the true answer should lie inside your chosen interval for 3 out of the following 5 questions: i.e. that the true outcome should lie between your "lower bound" and your "upper bound" 60 percent of the time. This is called a 60% confidence interval or 60% CI. "Easy" questions would have a narrow 60% CI while, in your opinion, "hard" questions may need a wide interval to reach the required 60% CI. By definition, your range for both types of questions should include the true outcome for the same 60 percent of the time in your opinion.

Please think about each question carefully and answer each based on what you know now. It is possible to be *over*confident or *under*confident and either type of mistake has a cost associated with it. If the true outcome lies inside the stated interval *more* often than 60 percent (i.e. the interval is too wide) then the advice was inappropriate because some risks were over-estimated and the business strategy cost money because it was poorly adapted to the true situation. If the true outcome lies inside the interval *less* often than 60 percent (i.e. the interval is too narrow) then the advice was inappropriate because some risks were underestimated and that misunderstanding cost money.

The five "P" questions ask you to estimate a probability for a given interval. To make your estimate of confidence clear, consider the following. If you say "P= 40%" to all of 5 questions then you are saying that you expect the true answer lies inside the stated range for 2 of the questions but outside the range for 3 of the questions (but you are not exactly sure which answers will be inside and which will be outside). If you say "You are 100% confident" to all questions then you expect the true answer to every question lies within the stated range. If you say "You are 0% confident" to all questions then you would be surprised if even one of the true answers was inside the stated range.

Q1 According to Statistics Canada, a certain number of people were living in Canada on July 1, 2009. Based on what you know now, what are upper and lower bounds for a 70% CI on that number?

Lower Bound _____ million Upper Bound _____ million

Q2 The Globe and Mail ranks Canadian companies annually according to their revenue and some companies earn a lot of revenue. Based on what you know now, what are upper and lower bounds for a 70% CI on the most revenue earned by a single company during 2008?

Lower Bound $_____ billion Upper Bound $_____ billion

Q3 Some people eat at restaurants often while other people need a food bank to survive. Based on what you know now, what are upper and lower bounds for a 70% CI on the fraction of Canadians who eat out at a restaurant at least two times per month?

Lower Bound _____ % Upper Bound _____ %

Q4 Facebook is a popular website which allows users to have "friends". Based on what you know now, what are upper and lower bounds for a 70% CI on the average number of friends per user?

Lower Bound _____ Upper Bound _____

Q5 Maclean's magazine is one of the most widely-read magazines in Canada. Suppose you had a budget of $250,000 to be used to buy full page, 4-colour nationally-distributed ads in Maclean's (e.g. no discounts, such as volume discounts). Based on what you know now, what are upper and lower bounds for a 70% CI on the number of ads that could be bought using this budget?

Lower Bound _____ Upper Bound _____

P1 World-wide exports are important both in terms of the dollar value and as a fraction of the economy. Based on what you know now, how confident are you that the fraction of Canada's Gross Domestic Product (GDP) in 2008 represented by exports was between 34% and 40%?

Confidence (P) _____%

P2 Based on what you know now, how confident are you that the number of square feet in a standard Canadian Football League football field (excluding endzones) lies between 55,000 sq. ft. and 72,000 sq. ft.?

Confidence (P) _____%

P3 Statistics Canada conducts surveys to learn how many households own their own home. Based on what you know now, how confident are you that the fraction of households who owned their own home in 2008 lay between 70% and 85%?

Confidence (P) _____%

P4 Social scientists study how people use their time. Based on what you know now, how confident are you that, according to Statistics Canada, the average amount of sleep for a

Canadian aged 15 to 24 lies between 425 and 510 minutes per day?

Confidence (P) _____%

P5 The University of Guelph is neither one of the biggest universities in Canada nor one of the smallest. Based on what you know now, how confident are you that the number of full-time equivalent students enrolled at the University of Guelph for 2008 lay between 15,500 and 19,300?

Confidence (P) _____%

Discussion

The correct answers to these questions can be found in this footnote.[22] If you got the correct answer precisely every time, congratulations and maybe you should apply for your boss's job. If you missed one answer, or got only one of them right, then you may want to learn more facts but your ability to answer a factual question precisely is not the point of this test.

The two styles of questions investigate different perspectives on the same issue. The first five questions ask you to provide a 60% CI. You should expect to get 3 out of 5 of them right but you may have got more or less. The following table reveals what it means. Suppose that the correct answer lies inside your range for 2 of the 5 questions. Since you can choose a wide or narrow interval for each question separately according to what you know, it does not really matter which two were "right". If your judgment is right but you were unlucky and only two of your estimates were "right" on these five questions then the vertical box in the table below shows common such luck is: aiming for 60% accuracy produces exactly 3/5 right only 35% of the time and produces exactly 2/5 23% of the time. On the other hand, if only two estimates were right in this test because your aim was not what you thought it was then the horizontal box in the table shows the relative probabilities[23] of the alternatives. For example, getting 2/5 right would occur with 23% probability if 60% accuracy but the evidence is stronger that the true accuracy is at 40%, since getting 2/5 right occurs with a 34.6% probability. Obviously, with 5 questions, luck plays a role but this table helps to give a sense of which kinds of differences are surprising and which differences are easy to rationalize as luck.

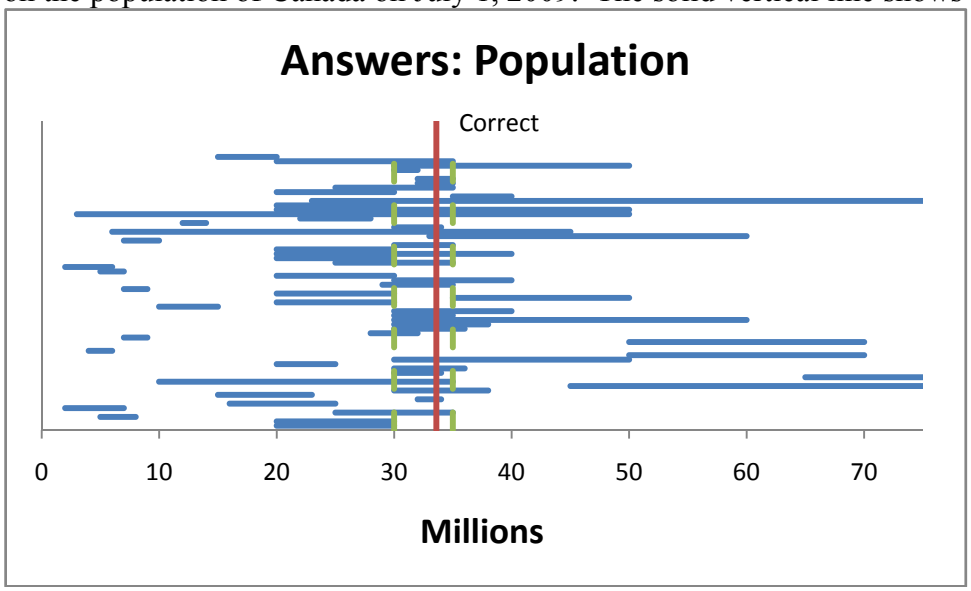
	Probability								
#Times Inside Interval (/5)	10%	20%	30%	40%	50%	60%	70%	80%	90%
0	59.0%	32.8%	16.8%	7.8%	3.1%	1.0%	0.2%	0.0%	0.0%
1	32.8%	41.0%	36.0%	25.9%	15.6%	7.7%	2.8%	0.6%	0.0%
2	7.3%	20.5%	30.9%	34.6%	31.3%	23.0%	13.2%	5.1%	0.8%
3	0.8%	5.1%	13.2%	23.0%	31.3%	34.6%	30.9%	20.5%	7.3%

[22] P1 Exports/GDP:35.6%; P2 CFL field size 66330 sq. ft.; P3 Homeownership Rate 65.9%; P4 Sleep 522 min/day; P5 Students 18300 Full-Time Equiv.; Q1 Population: 33.6m; Q2 Revenue: $37.5b; Q3 Restaurant meals/mth: 61%;Q4 Friends/User 120 (Facebook), 255 (another source; 18- 25 yr Cdns.); Q5 Ads 7.4

[23] This table uses probabilities computed using the Binomial distribution. If you want to be picky, this discussion of probability uses a "frequentist" perspective, which is familiar in a repeatable laboratory experiments and differs from the "subjective" perspective often used in business. I think that this frequentist perspective is reasonable in this context since the experiment can be repeated with more questions and modified questions.

4	0.0%	0.6%	2.8%	7.7%	15.6%	25.9%	36.0%	41.0%	32.8%
5	0.0%	0.0%	0.2%	1.0%	3.1%	7.8%	16.8%	32.8%	59.0%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

It may also help to see what others have done on similar questions. For question Q1 on Canada's population, the following graph shows the ranges given by 60 undergraduate students in an Introduction to Marketing class when asked for a 70 percent confidence interval (70% CI) on the population of Canada on July 1, 2009. The solid vertical line shows the right answer.



Despite being a number that is important and familiar in many contexts, including contexts relevant to students in an introductory marketing class, almost everybody offered a surprisingly large range for population. Many ranges did not include the correct answer. For example, this figure also indicates that many of the respondents would not be surprised if the search engine had said something stupid, such as "the population of Canada was 20 million". One can argue that these answers have no serious implications because anybody who needs to know the national population can find it quickly using a search engine like Google.

Awareness, more than the facts, become more important as you become more senior since it may not be possible to use Google for the kinds of questions which senior executives must answer. Search engines do not help when trying to ask the right question and knowing what question to ask is related to your awareness of the quality of your judgment.

The second type of question investigates judgment in this type of scenario since the question is more closely related to a question which asks how willing you are to bet on a given answer. In principle, since each of the questions has a specific answer, somebody could sincerely answer 0% or 100% to each of the answers and get a "perfect" score. Most people do not know the answers to all of the questions, and give a number in between. Ideally, the number should be very high if the correct answer lies in the given range and very low if the correct answer lies outside of the given range. After learning the correct answer, you should be surprised with yourself if you said 0% on a question where the correct answer was inside the given interval or if you said 100% on a question where the correct answer was outside the given

interval (and especially if the correct answer was far outside the interval you provided).

It is possible to rationalize why any one answer should be counted as "almost right", *after* learning its answer. The discussion in the main text shows that such an attitude is dangerous for several reasons. First, as noted in the initial motivation, some people advocate using judgment in a decision making process but, in practice, decisions are made before all of the facts are known. Second, since judgment is not a "magic key", it is evident in a pattern of knowledge: anybody can be lucky or unlucky once. If you have a good process to develop, test and refine your own judgment, you should not change your opinion of what these questions reveal just because somebody tells you the answer.

One can quibble with the test more intelligently: maybe the scoring should vary with the type of questions or maybe the score should allow some kind of "part marks" given for answers that were close. These quibbles are worth considering, but the various responses should also be obvious enough. It may also be worth noting that some research has shown how people and organizations fail to learn from answers that are almost right (e.g. Tetlock, 2007).

The first five questions on the confidence intervals are similar to the not-uncommon exercise of finding a "best case" and a "worst case" scenario. The second five questions on probability are also realistic in the sense of being linked to a risk-return trade off. These few questions have the advantage of being easy enough to write down in a format that can be read by many people at many different times. To a more select audience, it should be clear that the questions can change to increase the realism without changing the test of judgment (so long as there is a "right" answer which can be used to score the answers).

Appendix 2 A Personality Test
Which one of the following combinations of personality traits best describes you?

A- A fighting spirit- aggressive and courageous. A sensitive deep thinker who is true blue with friends.
B- Very intelligent and able to influence people. Good politician. Thirsty for knowledge. Talented and inventive, but easily discouraged.
C- Excitable, with complex life. Stubborn on outside, soft hearted on inside. A born leader.
D- Luckiest of signs- talented and affectionate, yet shy. You seek peace.
E- Shy but short tempered. Affectionate and kind to loved ones. You are impulsive and honest.
F- Elegant and creative- best in arts. Passionate but timid.
G- A pioneer spirit devoted to work and quest after knowledge. Sometimes selfish and eccentric.
H- Honest and loyal leader of men. Generous but stubborn. Earmarked for success.
I- Cheerful and popular, but impatient. Handy with money- always a winner.
J- Wise and intense, but vain. Passionate and determined. A winner with money.
K- Ambitious and honest but prone to spend freely. Quick to anger.
L- Bright, inspiring and easy going. You make an outstanding parent.

STOP

Choose one of the above before reading the discussion on the next page.

Discussion

To ensure the highest quality data source for this Test, the list of personality traits was taken from a placemat at a Chinese restaurant celebrating the Chinese New Year. Each year of the 12 year cycle of the Chinese Zodiac is named after a different animal that is supposed to impart distinct characteristics to people born in that year and that to determine that person's personality, physical and mental attributes, success and happiness throughout his/her lifetime.

Letter= Animal	Birth Year[24]
A = TIGER	1914, 1926, 1938, 1950, 1962, 1974, 1986, 1998, 2010
B = MONKEY	1920, 1932, 1944, 1956, 1968, 1980, 1992, 2004, 2016
C = DRAGON	1916, 1928, 1940, 1952, 1964, 1976, 1988, 2000, 2012
D = RABBIT	1915, 1927, 1939, 1951, 1963, 1975, 1987, 1999, 2011
E = PIG	1923, 1935, 1947, 1959, 1971, 1983, 1995, 2007, 2019
F = SHEEP	1919, 1931, 1943, 1955, 1967, 1979, 1991, 2003, 2015
G = ROOSTER	1921, 1933, 1945, 1957, 1969, 1981, 1993, 2005, 2017
H = DOG	1922, 1934, 1946, 1958, 1970, 1982, 1994, 2006, 2018
I = HORSE	1918, 1930, 1942, 1954, 1966, 1978, 1990, 2002, 2014
J = SNAKE	1917, 1929, 1941, 1953, 1965, 1977, 1989, 2001, 2013
K = RAT	1924, 1936, 1948, 1960, 1972, 1984, 1996, 2008, 2020
L = OX	1913, 1925, 1937, 1949, 1961, 1973, 1985, 1997, 2009

This test has a serious edge to it. Just as many people believe some strange things about their own business very strongly, many people believe in astrology very strongly. People may justify their belief to themselves by pointing to statements written for them, based on their birth date for example, and noting the surprising accuracy of such statements. Such justifications represent a confusion of prediction and post-diction, because the statements being judged tend to be vague and open to multiple interpretations. This test explores the meaning of "surprising" where self-deception is common: instead of evaluating the accuracy of a statement which is expected to be accurate, this Test asks you to select one from a much longer list of statements. The accurate statement should be easy to identify if the prediction is correct.

"The Amazing Randi" used a variation on this experiment and, because of his background and because of the more elaborate implementation, the variation offers a more effective lesson. (See http://www.youtube.com/watch?v=tFwOeXi3Yz8 for a video demonstration used in a PBS show). Taleb (2007) refers to a similar kind of misconnection in the context of an ice cube: starting with a frozen ice cube, it is easy to foresee what the puddle will look like after it melts. But starting with a puddle, it is much harder to see what the ice cube looked like.

The lesson from this test should be that it is hard to forecast when you do not know what will happen. This lesson also shows how to avoid the problem: be systematic and act as though you do not know the answer.

[24] There is a small detail that the Chinese New Year occurs in late January or early February of the year listed. But, to use an argument explored in greater depth elsewhere, this detail should result in about 1/12 of the responses being incorrectly categorized (but in a way that could be easily corrected for in a careful study). Therefore, the big question is whether the overall performance is sensitive to this 1-in-12 problem.

Appendix 3: Fermi Questions

Scientists helped move the sun to the centre of the solar system, identified the atom, created the hydrogen bomb, enabled the technology used to create and read this document and much else. Yet, many students and many with an interest in business are not *curious* about science in general or physics in particular. That attitude is unfortunate since physicists developed a style of questions called "Fermi Questions" or "Fermi Problems".

These questions stimulate students to apply what they know and, more importantly, to recognize how what they know fits together in a way that enables them to answer something more puzzling than a word problem designed to apply a textbook formula. Unlike the questions that might earn a top prize in a Jeopardy style quiz show, success in answering these questions is not based on memorizing a textbook, or even being able to search the Web really fast. Typical Fermi Questions include

What is the weight of the air over Lake Superior? (measured in Newtons)

When the island of Krakatoa was destroyed by a volcanic eruption, the sound waves could be detected world wide. How long would it take for such a wave to travel around the earth and come back to Krakatoa? (measured in seconds)

(Source: http://www.physics.uwo.ca/science_olympics/events/puzzles/fermi_questions.html)

It may help to note that a surprising number of the questions are business related. For example,

How many piano tuners are there in New York City?[25]

How many shopping malls are there in the US?

http://www.vendian.org/envelope/dir0/fermi_questions.html

http://scienceblogs.com/principles/2007/05/basic_concepts_fermi_questions.php

Weinstein and Adam's book (2008) offers a similar perspective on what it is important to know and offers many questions.

The strategy when working on these types of questions is to work out what you need to know and to estimate it. As one website notes, points can be given for the quality of the final estimate where quality is measured by "factors": being wrong by a factor of 3 implies that the ratio of the estimate to the true answer (or judge's idea of truth) is between 1/3 and 3 (http://www.physics.uwo.ca/science_olympics/events/puzzles/fermi_questions.html). Weinstein and Adam offer a similar perspective by repeatedly discussing the meaning of an "order of magnitude." In this way, students can learn both what is important and what is reasonable. There may also be more than one way to find a reasonable answer: Weinstein and Adam give some suggestions on how to approach a question using "bottom-up" reasoning and "top-down"

[25] http://mathforum.org/workshops/sum96/interdisc/classicfermi.html shows how this process works:
Approximately how many people are in New York City? About 10,000,000. [clearly this calculation is a bit old]
Does every individual own a piano? No. Would it be reasonable to assert that individuals don't tend to own pianos and that families do? Yes. About how many families are there in a city of 10 million people? Perhaps there are 2,000,000 families in NYC. Does every family own a piano? No. Perhaps one out of every five does. That would mean there are about 400,000 pianos in NYC.
How many piano tuners are needed for 400,000 pianos? Some people never get around to tuning their piano; some people tune their piano every month. If we assume that "on the average" every piano gets tuned once a year, then there are 400,000 "piano tunings" every year.
How many piano tunings can one piano tuner do? Let's assume that the average piano tuner can tune four pianos a day. Also assume that there are 200 working days per year. That means that every tuner can tune about 800 pianos per year.
So, how many piano tuners are needed? The number of tuners is approximately 400,000/800 or 500 piano tuners.

reasoning.

With this context, I would like to suggest some Fermi Questions that would be relevant to business students. I am not sure if it is possible to find the correct answer to any of these questions but I would be happy to add questions and answers if people wish to suggest them.

1/ How many chocolate bars are consumed in a year in Canada?

2/ If a loaf of Wonder Bread sells for $2 today, how long ago was its price one half of its current level? (Note this is easier than a forecasting exercise which would ask for a time to double, but uses the same skills.)

3/ What is the ratio of the number of bottles of Coca-Cola consumed in the US relative to the number consumed in rest of the world?

4/ What are the sales revenue per square foot at Walmart? (measured in dollars)

5/ How many men's shirts are sold in a year?

6/ At this university, how many people use a foodbank?

7/ On average in Canada, how many hours does a person need to work to pay for a new car? (Source, modified from *Guesstimation*)

8/ During the last big California earthquake, two million books fell off the shelves in a university library. How many students would need to be hired to reshelf all of these books in three weeks? (Source: *Guesstimation*)

9/ What is the risk of dying per mile traveled in a car?

10/ What fraction of deaths in Canada is caused by cars?

11/ On average, how much does each cigarette smoked by a heavy smoker shorten his or her life? (Source: *Guesstimation*)

12/ During the last 250 trading days (i.e. about a calendar year), how many days did the TSX change by more than 1 percent up or down?

13/ How many power plants are needed to power the future fleet of electric automobiles?

14/ How many words does a university student hear spoken by the instructors during a typical four-year degree program?

15/ According to Westjet, what percentage of Cdns owns credit card? 85%

16/ According to Westjet Airlines, how many complimentary snacks do WJ guests eat on a given day? 32000 cookies and snack mix

17/ Measured in terms of dollars, what is a 1 percent market share of the market in X?

18/ How many golf balls fit in a school bus?[26]

19/ How many words are spoken in an average TV advertisement? How many words are spoken in an average radio advertisement?

20/ How many cups of Tim Horton's coffee are sold in a year? (This question is not trivial: it impacts the use of paper by Tim Hortons Ltd., the recycling and environmental policies of local communities, and the cost of their R-R-Roll up the Rim contest.)

Additional Discussion

As these notes try to argue, the judgment and awareness invoked when answering a Fermi Question are vital business skills invoked before you decide whether it is worthwhile spending the time, money and effort needed to carefully evaluate a situation as part of a business plan. More accurate answers are better but, if you get good at answering certain types of

[26] *Report on Business Magazine*, Oct. 2010, p. 48; answer= about 500,000. Would it make much difference if "ping pong balls" replaced "golf balls"?

questions in business, then that fact is often taken as a sign that you are ready for a promotion where the questions get harder and less raw data is available for decision-making.

Arguing that these questions are not precisely defined (e.g. in question 3/, are we talking about 500 ml bottle of Coca-Cola or a 333 ml can or 250 ml? In question 4/, Should we adjust prices for inflation) should probably be penalized when the concerns are insignificant. Such definitional issues should be clarified by stating the thought process precisely so that, if there were a dispute, the effect of differences could be debated. Other issues may have a bigger effect on the final answer and should be reasonably known to a serious student. Notice also that the ratio of 500/333 represents a factor of only 1.5. These questions may also help to show why math skills are useful, since the estimation process offers a built-in measure of the relative significance of different variables, and to show that some uncomplicated bits of math are *very* important.[27]

An interesting feature of this style of question is that, after doing a couple of questions where the instructor knows the correct answer, a student can introduce a question where the instructor does not know the answer. To the other students, the process of argumentation and estimation would not change. Even if other students had access to insider information about the answer as a fact, they may not be able to justify their estimate without revealing that they knew the answer.

[27] Since the scoring is based on factors, orders of magnitude or other measures of proportional difference, it may help to note the contribution of different kinds of errors in the bits to the error in a combined estimate. For example, consider a simple question where Sales= Population* Q and Q represents the consumption of a typical consumer. While it might be nice to both Population and Q precisely, it is also approximately true that

%d Sales= %d Population+ %d Q

where %d… shows the percentage difference. Therefore, if the estimate of population is too high by 5 percent and the estimate of Q is too low by 10 percent then the resulting estimate will be too low by 5 percent.