Thoughts on Diagnostic Accuracy and Spiritual Discernment

Consider all events and motivations to be of two types: those that derive directly from God's actions in the world and in the life of the believer and those that do not. Events and motivations falling in the latter category could be those related to (i) the consequences of our own or others' sinfulness, (ii) events that are random (choose your own definition), or (iii) choices we and others make that are within God's permissive (*vs* prescriptive) will. Some may argue the existence of (ii) and (iii) or their separability from the first category, but we can probably agree that the existence of (i) is sufficient reason to ponder the issue.

It seems to me that, through the process of sanctification, the Spirit brings about an increasing level of discernment within the Christian with respect to distinguishing these two categories. Consider Figure 1 below, which represents events that are truly attributable to divine agency with red bars and events that fall into (i), (ii), and (iii) above with blue bars. Unless one believes in entire sanctification (not I), we will never, even with the help of scripture, wise counsel, and Christian community, eliminate the overlap of the red and blue distributions due to the noise of our own sinfulness. The new/immature Christian in the upper left may tread on the thinnest ice, whereas the most spirit-led among us may come close to the lower right. Andy may take issue with the degree to which anyone can approach the lower right in this lifetime. Others might argue that the total number of events represented by the blue/red bars should not be equal as depicted below.

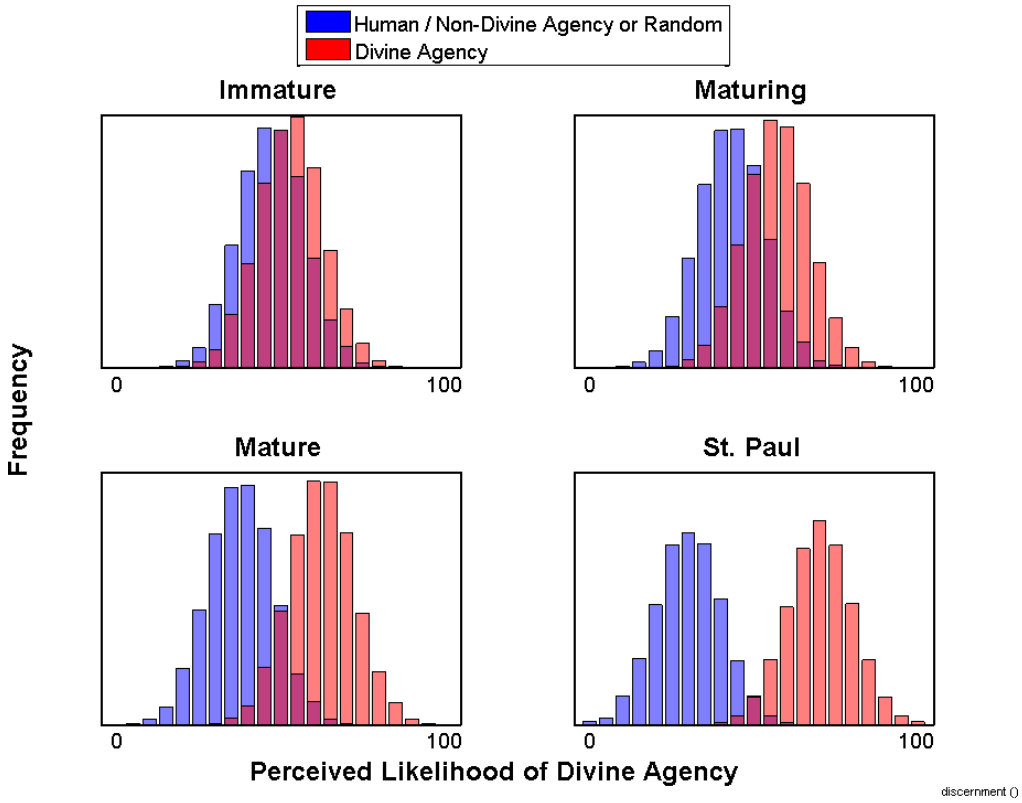

Figure 1. Notional representation of Christians' ability to discern events and motivations having divine (red bars) or other (blue bars) origins.

Because of the overlap in these distributions, we must make the best possible informed choices while facing the reality that errors are inevitable. Generally, there are three types of choices one can make:

1. ***Choose not to attribute.*** This decision can be conscious or unconscious. One might choose not to classify an event that is perceived to be trivial or insignificant (*eg*, the change in direction of a dust particle suspended in the air). Even if an event is not deemed trivial or insignificant, one might choose not to classify because of the perception that an error in classification has a trivial or insignificant consequences. One might also choose not to classify due to a lack of spiritual discipline or out of spiritual indifference.

2. ***Attribute an event or motivation to divine agency.*** Again, we may not be fully conscious of this decision.

3. ***Attribute an event or motivation to non-divine agency or chance.***

If a decision is made to make an attribution, one must further decide, implicitly or explicitly, what threshold to apply for classifying any given event. Consider the "Mature Christian" depicted in the lower left subplot in Figure 1, reproduced as Figure 2 below.

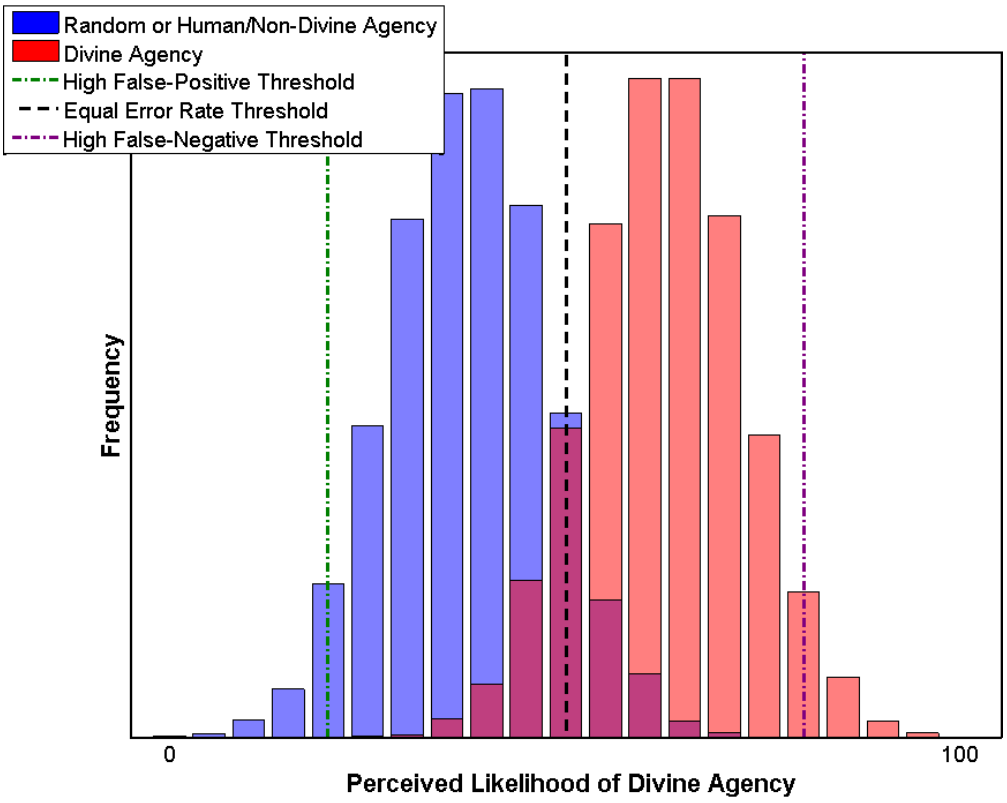

Figure 2. A subplot from Figure 1 above with thresholds corresponding to highly sensitive (green dash-dot line) and highly specific (purple dash-dot line) decision-making. The black dashed line represents a decision threshold that would yield an equal number of false-positive and false-negative attributions.

One might suggest that the optimum location for a decision threshold would be that depicted by the black dashed line. This decision threshold makes the proportions of the two types of possible errors equal: one is just as likely to incorrectly attribute God's actions to chance or non-divine origins (a False Negative, FN, error) as to incorrectly attribute events or motivations that originate from random chance or human agency to God (a False Positive, FP, error). However, certain decisions or conditions may call for erring toward thresholds that generate a larger proportion of FN or FP attributions.

For example, in the now-classic case of attributing the origins of an optimally-located parking space, one may deem the cost of a FP error to be small, while the cost of a FN error might be the unintended cultivation of a tendency to avoid seeking out and appreciating God's providence. In this case a low decision threshold (the green dashed line) might be judged most appropriate. Along with this threshold comes the knowledge that FP errors will be made and, hopefully, a sensitivity to the ramifications of these errors in one's own life (esteeming oneself too highly, perhaps) and the lives of others (for example, the risk that such an intentional bias may affect one's witness to skeptics). Other rationales are, of course, entirely possible. I will happily dodge the issue of whether one ought to regularly petition the Almighty for such consideration.

On the other hand, consider the case of the minister who is confident that God has given him a sign that his congregation ought to undertake the construction of a new sanctuary, and that the means for financing this effort will be provided in an exceptionally short amount of time. The cost of a FP attribution may be extremely high in this case. Thus, a high decision threshold might be chosen (purple line), and the requirements and steps one might take prior to making such an attribution might be greater in number and more exacting in rigor. There is a cost to this type of threshold, as well: it runs the risk of missing out on the great blessings associated with exercising our faith and relying upon God rather than our own strength.

One can compute the percentage of FP errors (labeled "False Positive Fraction" below) and the percentage of events and motivations correctly classified as having a divine origin (classically and spiritually, the "Sensitivity") for every possible decision threshold. Plotting these two parameters against one another yields what is commonly called a Receiver Operating Characteristics (ROC) curve. ROC curves were originally developed to characterize the range of tuning parameters used to optimize a radio receiver's ability to detect signals in the presence of noise under varying battlefield conditions: a metaphor that is not altogether inappropriate. The ROC curves for our friends depicted in Figure 1 are shown below in Figure 3.

Better discriminatory power (or diagnostic accuracy) is represented by ROC curves that more closely approach the upper left corner of the plot. Attributions that are 100% accurate achieve this point. The opposite extreme, attributions that are completely random and uninformed, is represented by the dashed black diagonal line that goes from the lower left corner to the upper right corner of the plot.

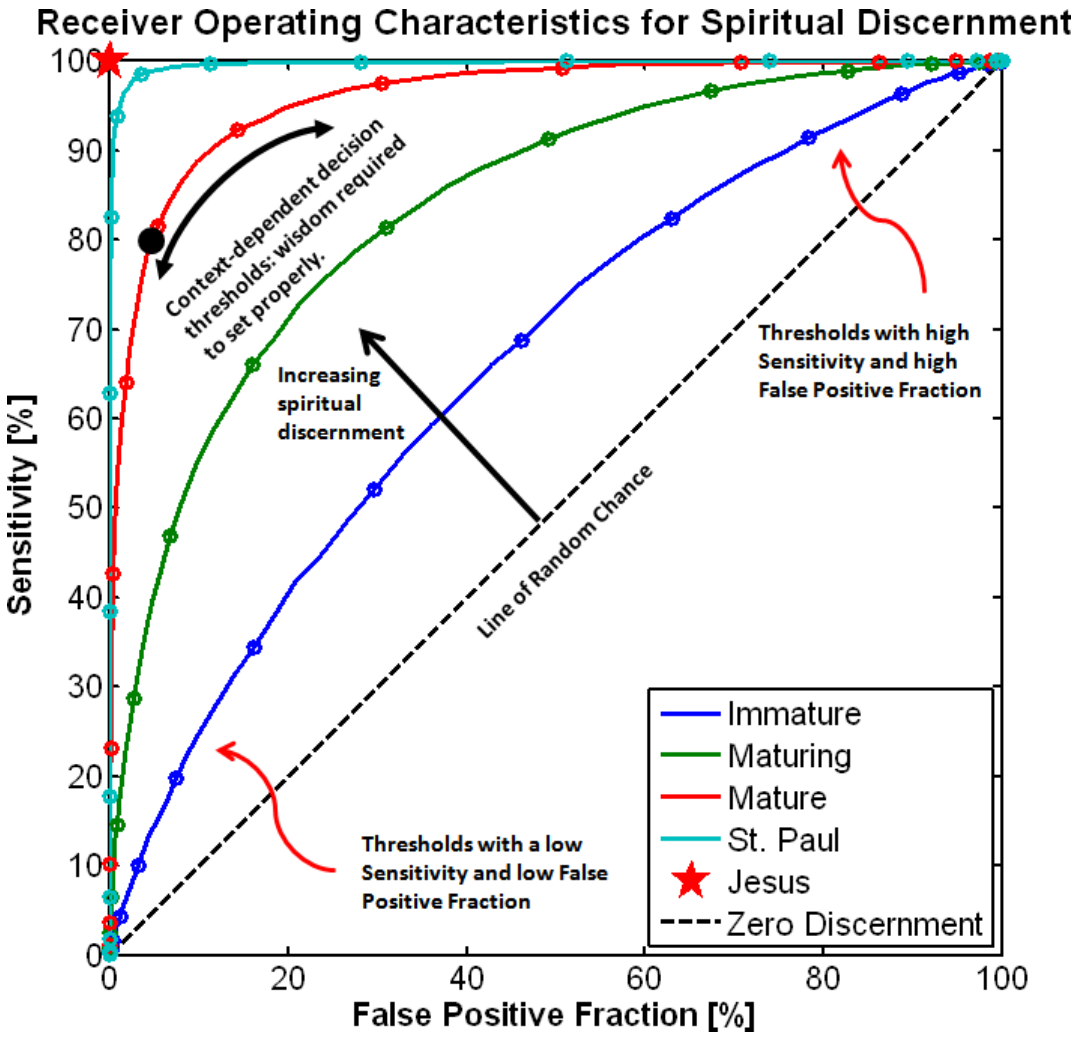

Figure 3. Receiver Operating Characteristics (ROC) curves for the distributions depicted in Figure 1.

Moving to a better ROC curve: The attributional accuracy of an individual or group of Christians who seek to discern the Spirit's leading from the noise of sin and the distractions of this life ought to improve with sanctification and the developmental processes, discussed in prior meetings, that set apart our hearts and minds to God. Of particular value is the counsel of spiritually mature individuals within one's Christian community. To a degree, consulting with a variety of individuals ought to average out some of the errors to which we are all prone[1]. We must remain aware, however, that errors are only averageable if they are random, and, as we studied last week, sources of systematic error abound. Some of these are inherent to our human condition and can only be conquered through God's grace. Another potentially-dangerous class of systematic errors may be those associated with one's particular denomination or faith tradition.

[1] For waging war you need guidance, and for victory many advisers. Proverbs 24:6 (NIV).

The costs of misclassification or non-classification: As discussed last week and described above, any attribution comes with the risk of error, and each type of error typically carries a penalty of some kind. Some penalties are severe while others may be trivial. It seems likely that there can also be a risk or penalty associated with the choice to not attribute. My guess would be that most of us make the latter choice, albeit passively, hundreds of times a day; to not do so might be crippling. We have reflected as a group on the cost of false-negative and false-positive attributions. It might also be worthwhile to consider the cost of failing to summon the fortitude to decide.

Discernment as an exercise in experimental design: When researchers conduct experiments, they are frequently designed to test a null hypothesis. Often, a positive outcome is associated with rejecting the null hypothesis in favor of some alternative, and a negative (null) outcome is associated with failing to reject the null hypothesis.

A well-designed experiment must simultaneously minimize the risk of two different types of errors. A Type I error (or *alpha* error) occurs when the null hypothesis is rejected wrongly, that is, when the experimental data yield a positive outcome despite the fact that the null hypothesis should have been retained. Type I errors are therefore most commonly false positive outcomes. (The estimated Type I error rate is the value typically specified in most experiment reports: a *p value* of 0.05 means there is just a 5% probability that a false positive conclusion has been reached.) A Type II error (or *beta* error) occurs if the experimenter fails to reject the null hypothesis when, in truth, it should have been rejected. Type II errors are therefore a type of false negative outcome. Although less commonly reported, the estimated Type II error rate is determined by the *statistical power* of the experiment: an experiment with a power of 80% has been designed such that the estimated likelihood of a false negative conclusion is 20%.

The optimal experimental design depends on the details of each situation, but it is relatively rare to find experiments that are designed to have an estimated alpha Type I error rate of more than 5% or an estimated Type II error rate of more than 20% (or, in other words, 80% sensitivity). These conditions are indicated by the black point on the red “Mature Christian” ROC curve in Figure 3. It is somewhat sobering to consider the discriminatory power associated with this experiment that, from a statistical perspective, has the minimum amount of discriminatory power to be considered worth doing. In this context, it might indeed be said that, when it comes to spiritual discernment, many or most of us are “experiments gone awry.”

Since we are surrounded by such a cloud of poor attributions: Last week we discussed some of the attributional biases to which we are prone, and we focused on a few to which people of faith may be particularly susceptible. We also discussed the problem of FP errors with respect to our witness to unbelievers. (“I’m not a Christian because I have so little respect for all of the Christians I see who attribute every small, random, happy event in their lives to God’s smiling upon them.”) It is undoubtedly true that such errors are made. In fact, Christians as a group may, intentionally or otherwise, set thresholds that result in a large number of FP errors. As described above, such a threshold may not be altogether unreasonable, but it still leaves open the question of how to respond to the skeptics and seekers we encounter in our lives.

Last Friday, I believe David and others alluded to the following response that I might try on a fellow nerd: You've chosen to engage in the field of (physics, engineering, psychology, *etc*). And yet, all around you, you could probably identify instances of poor methodology and decision-making: people who seem to willfully blind themselves to patterns in data that don't fit their hypotheses and biases, who seem to apply an interpretive bias to their observations far more often than they ought, who design flawed experiments, and who request and receive funding for research efforts you believe to have questionable value. In your more honest moments, you might even admit to being prone to such errors yourself. Still, you continue to strive for an ideal that is likely unattainable. You work in a sea of people (none at your own institution, of course) who are poor performers and who operate with questionable discriminatory power. Yet you continue to believe that your work has value and that, through the scientific community, your discipline continues to advance in meaningful ways. You might dismiss the possibility of God's existence and/or His interest in your life, but it is probably disingenuous to dismiss God based on the reality of the many errors made on the part of those who follow Him.