Chapter 10

Manager Selection

Designing a Manager Selection Algorithm

Is Warren Buffett a good investment manager? It seems like such a seemingly simple question, but is it really? How do you define *good*? What factors should you consider? Is his entire track record relevant? The past 10 years? Five years? Last year? Should we consider only his average annualized returns? Calendar years or rolling 12 months? What about his drawdowns? Volatility?

Would your process differ if you were attempting to assess a hot-shot trader from one of the big investment banks who doesn't have a track record as a CIO or even as a portfolio manager? How about a manager who's been generating exceptional returns on a small portfolio out of his garage, but has never worked at a large institution? Does it make sense to fast-track one, but not another?

These are the questions that allocators should grapple with long before they invest a penny, but few do. Instead, fund managers are asked to fill out due-diligence questionnaires with 400 or more questions. Consultants produce extensive reports and conduct intensive on-site assessments. For what? Just as most portfolio managers never truly

AlphaBrain: How a Group of Iconoclasts Are Using Cognitive Science to Advance the Business of Alpha Generation, First Edition. Stephen Duneier.

contemplate the impact that sizing has on their investment returns, most allocators have done very little to understand the role that experience, assets under management, research techniques, personal investment, or any of the other factors that they diligently gather information about have on a manager's returns.

For all the meetings, questionnaires, consultants, and *expertise* employed, you would be hard pressed to find an allocator who consistently generates alpha, and a great many often find it difficult to even keep up with beta. Rather than admit that the problem lies in their analysis, the prevailing theme is that fund managers can't generate alpha consistently, therefore, the objective should be to generate beta at the lowest cost possible.

When people give up because a task requires cognitive strain in order to be done properly, it often means there is an opportunity to capitalize for those willing to put in the effort. We believe this to be the case here. So when a few large allocators came to us looking for help with manager selection we used it as an opportunity to build a new manager selection process from scratch. As always, we applied the fundamental principles of normative decision theory to the problem while using decision architecture to help allocators avoid the predictable cognitive land mines. The result is a truly unique approach that allows for an *objective* assessment that can be *consistently* applied to managers across asset classes and investment styles while also allowing for freedom of expression. I highlighted *objective* and *consistent* because those are the two key aspects that allocators and consultants almost universally lack, but without it the analysis has little value whatsoever, no matter how many questions are asked and on-site interviews conducted. I'll come back to the reason freedom of expression (i.e. beliefs) is not only valid in this process, it is actually necessary.

We began by asking our clients a simple question. "What questions do you need answers to in order to properly assess a manager?" With guidance from us along the way, after weeks of internal discussions, they honed in on their specific questions. For example, they wanted to know how much of their net worth the managers had invested in their own fund, the ratio of fee income to expenses and how many years of experience the CIO has. We then examined the purpose of each of those

questions to understand *why* they needed the answers to each of them. What value did they deliver? What would they do with that information and how would they go about factoring it into an actual decision?

We then asked them to weight the value of each individual question (variable) where the total sums to 100%. We also bucketed the questions according to three key factors: business risk, investment risk, and investment opportunity, as well as subsectors within them, and asked them to apportion weights accordingly. Again, they must add up to 100%. Interestingly, the results were different between the two queries. In other words, the very same variables were weighted differently, depending on how the we framed the query. That's a red flag, but a highly predictable one given what we know from the cognitive sciences.

A five-point scoring system was devised for each question, which was then weighted and summed in order to deliver a grade for the factor categories and variables.

It seems so simple, but as they say, the devil is in the details. For example, consider the following question under *Investment Risk -> Management Quality -> Experience of Senior Investment Team*. It could be asked as simply as:

How experienced is the senior investment team?
A. Extremely experienced.
B. Very experienced.
C. Experienced.
D. Somewhat experienced.
E. Inexperienced.

However, this range of options leaves a great deal up to the analyst's interpretation, and that means there could be great inconsistency even between funds being considered by a single analyst. One way to resolve that is to have these options:

A. More than 20 years combined in investment industry.
B. 15–20 years combined in investment industry.
C. 10–14 years combined in investment industry.
D. 5–9 years combined in investment industry.
E. 0–4 years combined in investment industry.

It's an improvement, but of course not all experience is the same. Perhaps we should break this one question down into two.

Combined years as buy-side risk-taker/decision-maker?
A. More than 20 years.
B. 15–20 years.
C. 10–14 years.
D. 5–9 years.
E. 0–4 years.
Combined years as sell-side risk-taker/decision-maker?
A. More than 20 years.
B. 15–20 years.
C. 10–14 years.
D. 5–9 years.
E. 0–4 years.

However, anyone who has ever managed a group of portfolio managers and overseen risk for a fund knows there is a difference between that job and managing a trading desk at an investment bank, and certainly a difference between it and managing a portfolio. So perhaps it would be appropriate to include the following questions and answers, first for the buy-side and then for the sell-side.

Combined years with directly attributable track record?
A. More than 20 years.
B. 15–20 years.
C. 10–14 years.
D. 5–9 years.
E. 0–4 years.
Combined years managing risk-takers?
A. More than 20 years.
B. 15–20 years.
C. 10–14 years.
D. 5–9 years.
E. 0–4 years.

You can see how it is that these questionnaires grow to be over 400 questions. You can also see how easy it is to get so bogged down in

the minutiae that you lose sight of how each of these details contributes to the overall purpose and the final decision. Eventually, the entire process becomes little more than a collection of *stuff* haphazardly collected and then applied according to intuition and gut feel. The output is riddled with cognitive bias, and offers little value beyond simply ticking boxes and covering behinds.

Flipping the Frame

There's one indisputable fact that must be acknowledged up front. We don't *know* how the number of years of experience as a fund manager contributes to returns. We don't *know* how the proportion of a CIO's wealth invested in their fund correlates with the returns of that fund. We don't *know* the relationship between the returns of a fund and just about all the answers to those 400-plus questions we love to ask with a purpose. The reason is that we are so confident in the value of our intuition and our assessment of what is "logical," that we don't see the need to gather evidence to support it. So here we are, decades into this process and we have little to no evidence to support the great majority of the allocation decisions being made. Trillions of dollars being allocated on little more than gut feel, and the results stand as proof.

In properly defining the decision problem we are attempting to solve we must acknowledge this as a "state." In other words, it is a factor that affects our ability to achieve the goal of allocating to better performing managers, but one we do not control … yet. Yes, although we don't yet know what factors are important, and to what degree, one of our key objectives is to ascertain that information. This is why we must allow the allocators to incorporate their beliefs, for those beliefs serve as their hypotheses. In order to prove or disprove their hypotheses they must apply them, observe them, and collect their findings. For only then can we move from the age-old belief-based system of manager selection to an evidence-based one.

As it is for all scientific inquiry, it is vital that we create a controlled experiment, using a repeatable and consistent process each and every time in order to collect the data. As the findings are gathered we can then

learn, make adjustments, and improve the process. In developing this controlled experiment we need to allow for the freedom of hypothesis development within the rigor of an objective, systematic process.

There's one additional factor that is fundamental to the objective of the experiment, and it relates to the nature of the query itself. What we seek is improvement to generate better results. As it is for all improvement, gains are achieved by reducing mistakes. In other words, gains are reductive, not additive. This applies to the manager selection process as well. Rather than attempting to identify the next stand-out manager who will produce upside outlier returns (a low probability event), we seek to predict the probability that a manager will make mistakes (a high probability event), and discount the expected returns accordingly. When you reframe the decision problem in this way, and then revisit those 400-plus due-diligence questions, it will become obvious that is exactly what they are meant to deliver. They help us understand who will underperform due to mistakes, and why.

Consider the question, "What is the ratio of fee income to expenses?" Why would we ask that? The reason is that if the fee income doesn't cover expenses there is a greater chance that the partners and employees will be concerned about the long-term prospects of the firm. That concern is a distraction from the task we want everyone's attention to be focused on – namely, the task of generating investment returns. So, if we begin with a baseline expectation for returns, and that expectation assumes 100% of the investment team's attention will be focused on generating investment returns, potential distractions should be a reason to discount those expectations. As it relates to this particular question, the allocator must define how low they believe the ratio could go without becoming a distraction.

If a variable does not create a distraction, the baseline expected return should be unaffected. Let me restate that for emphasis. No distraction means no impact on expected returns. In other words, just because my expenses are covered several times over doesn't mean you should expect me to generate returns greater than my benchmark. However, if we're barely breaking even it could be cause for discounting those expectations.

Every question we ask is designed to help us discover whether a manager is more vulnerable to making a mistake. That's it. Just identify the probability of making a particular type of mistake and then apply

an appropriate discount to the baseline expectation in exactly the same proportion for every manager who exhibits the same shortcoming. It works for everything, from years of experience to the quality of financing partners.

Now that we understand the function of the questions and answers, we must determine an appropriate baseline return expectation. As it is for almost every entrepreneur, when you ask a fund manager what the appropriate benchmark is for their style, most will say their approach is so unique that it's hard to find a comparable benchmark. Nonsense. Even if there isn't a single index that is directly comparable, there is certainly some combination of indices that will allow for the creation of a synthetic benchmark.

Then the question is, what is the appropriate historical time period from which to set your expectations for the future? 30 years? 10 years? 5 years? 1 year? Do you weight each year equally? What about for a manager that has a track record? Should you completely replace the synthetic benchmark with the manager's returns? If you do, you are effectively saying that their returns are an absolute function of skill with no contribution from luck. Is that really what you believe? If not, then you should weight the synthetic benchmark and manager track record accordingly, but whatever your weightings, time horizon, and weighting of luck to skill, everything must be applied consistently to all managers.

We call this approach the Bija Manager Discounting Method. We've described it using broad strokes in order to focus on the fundamental difference between this approach and the industry standard. As you can imagine, there are a number of additional, very important components I have not mentioned, including levers for risk aversion, macro expectations, and other adjustments. The discounting method is but one component of an objective, evidence-based, and consistent investment process for allocators that enables improvement over time.

Stroke of Genius

Golfers like to say that the game is one of inches, millimeters even. The slightest of adjustments can yield dramatically different results. Take our eye off the ball or dip our head, and we're likely to send it on the

wrong trajectory, adding strokes to the final outcome. Performance is so finely tuned that etiquette dictates you should never talk during someone's backswing, it's bad manners to allow your shadow to cross between another player's ball and the cup while they putt, and sand traps must be raked smooth after use.

In truth, so many factors affect performance. There are the broad, static physical characteristics, external to the golfer, whose details are typically provided on scorecards and signage throughout the course. They include distance from tee to hole, shape and width of the fairway, placement, shape, and size of hazards, and even pin placement. Although they are the most commonly considered physical factors, there are plenty more that can have a significant effect on the success of an individual shot. Temperature, humidity, weather, and speed of play are just a few of the other factors external to us that can have an impact on our scores. When it comes to improvement though, little thought goes into the environment.

Instead, we tend to focus on what we believe we can control – namely, our bodies and the equipment that serves as an extension of us. Bigger, lighter, more spring, larger sweet spot, more comfortable grip, better feel, and that's just the club. Shirts that let your skin breathe. Balls with dimples that are scientifically designed to capitalize on microscopic changes in course conditions, and pressurized to be finely tuned to our swing. Shoes, gloves, hats, sunglasses, bags – there is no end to the ways we can purchase a better golf score. Considering the more than $4 billion per year that Americans alone are spending on golf equipment it's clear that improvement is desired.

Yet, with all the advancements in technology over the past 15 years and the insatiable desire among players to employ it, the average golfer's handicap hasn't shown improvement. They still require roughly 100 shots to complete 18 holes. Surely, in the hands of highly trained professionals, scores must have collapsed, right? Actually, they haven't really seen improvement either. In fact, you could argue that technology has actually stifled the impressive decades-long trend of improvement among professional golfers (see Figure 10.1).

I've heard all the reasons that the true impact of improved technology has yet to be reflected in golf scores, but the most common among them

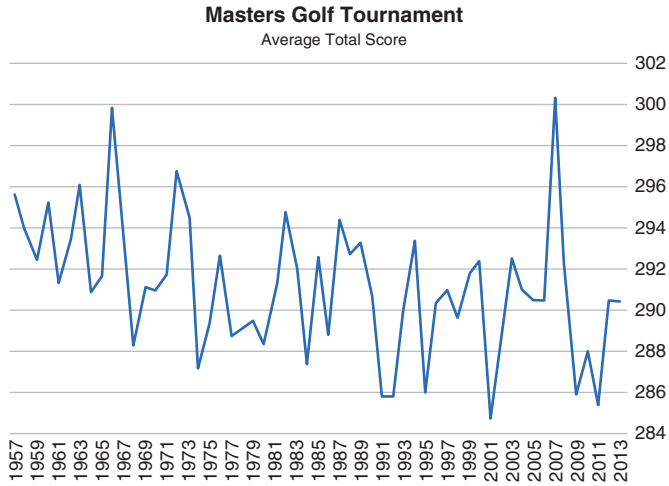

Figure 10.1 Masters golf tournament.
SOURCE: Bija Advisors.

is that they've made the courses longer, specifically to offset the effect of the new clubs. The evidence, however, does not support this argument. I gathered data from the Masters Tournament in order to weed out the noise and what better place to do that than Augusta. Rather than use winning scores (a far less time-consuming task), I compiled the average total score for all golfers who made the cut, so as to focus on the generic trend. If the injection of titanium heads were a major factor, we should be able to point to the year that it happened without any hesitation. If the delayed extension of the course's length had a great impact, we should likewise be able to point to the year the scores clearly reversed trend. Can you guess when technology transformed the golf club and the year when Augusta lengthened by 5% just by looking at the chart? (Hint: If you think you can you're mistaken.)

Turns out, in order to explain the outlier spikes and collapses in scores, you'd be better served looking to the weatherman, than the golfer's bag. Apparently, rain delays are a score killer and clear skies are like manna from heaven.

It's funny really. Anyone who has ever set foot on a golf course knows full well that it is a game of inches not yards. Yet, for some reason,

we still believe a club that adds yards to our game, and time spent at the range or with a pro focused on the distance aspect is where we will find some dramatic improvement. There are so many ways we can improve our game. For me, nothing comes close to the simple act of proper club selection. I don't mean knowing the club Rory would use in a particular instance, or even the one I or someone who scores similarly to me would use, but the one that fits my natural swing, flaws and all. When I have absolute confidence that I have the correct tool in my hand for the task ahead I swing more naturally, fluidly, and accurately. I am relaxed, and can focus on the other aspects of the game where my attention is better served.

For years I had two golf games. In one, I was all over the place. My club selection was haphazard, and every swing came with uncertainty. It was the kind of uncertainty that lead me to hold back or push harder at the very last nano-second of my swing. My scores were an accurate reflection of that confusion. In my other golf game, I was confident and inspired. My scores were consistent, and on average, 15 to 25 shots lower. The difference? I played with an experienced golfer who understood my swing. He never commented on my tendencies or suggested improvements. Neither did he offer me the latest weapon in his bag. He simply knew who I was as a golfer, and made my club selections for me. I had absolute confidence in his selections and so those misgivings that would normally tweak my swing, sending me off into hazards, disappeared. For those who have seen the movie, *The Legend of Bagger Vance,* you will understand when I say, it allowed me to "get out of my own way." When I golfed with this friend, I could "see the field," and my scores reflected it.

Unfortunately, Andrew King, my friend and guru of club selection, died in the World Trade Towers. I haven't experienced the game in the same way since.

The Definition of Actionable Intelligence

In many ways the game of golf serves as a metaphor for business and investing. Very often we spend the great majority of our time and effort

focused on the aspects of our job that provide the smallest potential for improvement while ignoring those that are far more likely to generate significant gains. Hours a day are spent talking to others, reading reports, and agonizing over the latest data, all in the hope of gleaning some insight that will improve our returns.

Trade ideas and market analysis are to investors what golf clubs are to golfers. You can feel them in your hand. You can hear that "ping" as the ball connects and the idea resonates. They are the go-to for investment performance improvement. I get it. I also get that the connection is more about convenience and availability than reality. The big gains come by analyzing our own tendencies, strengths, and weaknesses, individually and collectively as humans. The greatest improvement gains come when we can get out of our own way, and see the world more clearly.

As mentioned earlier, at one point I was asked to take over the investment process for a hedge fund. Considering the individual and fund performance until that moment, you might think my first points of order would involve tightening up risk parameters, improving our forecasting by hiring new analysts or subscribing to more research, or organizing more meetings in which we could share the market intelligence each PM was gathering individually. It's a natural assumption. I mean, if we could just improve our assessment of the markets we would improve our returns. We needed to get focused. Get down to brass tacks. Spend more time pouring over data, on research calls, in meetings, and of course, tighten up risk parameters.

I elected to do none of that. Rather than suggesting a club suited for Rory McIlroy, I spent time understanding each individual's natural tendencies, as well as those of the group, and then gave them the tools to improve their own individual games. I presented evidence for my findings and with each day, as returns improved, they began trusting my club selection, freeing them to swing more naturally, and their scores reflected it.

Through *AlphaBrain*, I'm not handing you the hottest new driver in my bag or even necessarily selecting the right club for you to use today. Instead, my goal is to help you make better club selections for yourself. That is how I define *Actionable Intelligence*.

What Allocators Can Learn From Paper Traders

Why is it that results generated through paper trading, meaning the trades aren't actually executed, are discounted relative to those that are actually generated? If you think about it, in both cases research is done, views are formed, expectations are set, instruments are selected, entry levels are determined, and decisions are made regarding trade exit. Doesn't that cover what almost every institutional investor spends just about every moment focused on?

Of course, for anyone who has actually managed capital, the difference is obvious. What the paper trader's experience lacks is the emotional impact of gains, losses, regret, and accountability. When real money is at stake repercussions from a misstep can involve serious consequences, and that effects the decision-making process. *That* is what distinguishes the results generated through paper trading versus actual investment management.

The fact that paper trading returns are discounted, rather than the other way around, implies that it is more difficult to generate returns when dealing with those emotional factors. I'd be hard pressed to find anyone who disagrees with anything I've said so far, which begs a few questions that surprisingly are rarely asked. If it is easier to generate returns when emotion is removed, why is so little time and effort spent focused on doing exactly that, and why do so many widely accepted, fundamental tenets of this industry serve to inject emotion into the decision making process rather than remove it?

What we are talking about here are decisions made with "affect-rich" outcomes versus those made with "affect-poor" outcomes. When risky decisions involve outcomes that can conjure considerable emotional reactions they are considered affect-rich. When they don't, they are categorized as affect-poor. It turns out, when we face decisions that have a greater potential to effect us emotionally, particularly those invoking negative feelings, we more often rely on our intuition and other mental shortcuts when considering our options while disregarding probabilities. In other words, our decisions become more emotionally driven and less probabilistic in nature which is exactly the opposite of what you want when attempting to improve the odds of a successful decision.

Although dozens of studies dating back to the 1990s have focused on and proven this case, one conducted recently took it one step further. For the first time researchers showed the existence of systematic preference reversals between affect-rich and affect-poor choices *within* individuals. In other words they showed that individuals who faced a problem with an affect-rich outcome would come to the polar opposite conclusion to the very same problem when it was later attached to an affect-poor outcome, and vice versa. This study – conducted in Germany and Switzerland – proved exactly what we've all suspected. There is a difference between managing a paper portfolio and managing a real one. These researchers proved that even the very same person, with the very same skill set, analytical tools, views, and expectations, when facing the very same decision, will make the opposite choices when emotions take over.

Think about that for a moment. When real money is on the line we are more likely to make *worse* decisions, and we know it! Now, think about some of the things we do on a regular basis that actually inject emotion into the decision-making process of those who are managing our money. Take, for instance, the common practice of requiring a CIO and other hedge fund partners to invest a substantial portion of their personal wealth in their fund. What is the rationale for such a requirement? To ensure that their goals and those of the investor are aligned. Well, if the investor's goal is for the leaders of that hedge fund to make decisions that are more probabilistic in nature, and less emotionally driven, thereby improving the odds of better returns relative to risk taken, the best way to align that goal with that of the manager is to *not* force them to invest a such a large portion of their wealth that is likely to turn an affect-poor decision into an affect-rich one. Ironically, that is *exactly* what the investors should hope to avoid but is precisely what they are accomplishing.

When I make this argument, particularly to asset allocators, I get more than a little pushback. Some will argue that it doesn't effect the managers, that they don't become more risk averse, which, if you think about it, is an odd argument to make. After all, isn't that precisely the intention, to keep the manager from behaving like a cowboy with your money? So, if you don't believe it actually makes the manager more risk averse, what exactly does it mean to be "aligned"? Research

shows that when you convert decisions from affect-poor into affect-rich ones, you don't only simply make choices more risk averse. As I will discuss next, there are times when affect-rich decisions will have the opposite affect. The only true constant is that the decisions lack the kind of objectivity and probabilistic foundation you want from someone managing your money.