The Persistence project: Prolegomena to industry monographs
**Motivations**

The Persistence Project is an attempt to identify the management behaviors that cause long-term superior performance. In other words, we’re trying to find the secrets of success.

To uncover these secrets, we have identified three firms in each of a dozen industries, each firm with a specific profile: “Miracle Worker” (MW), “Long Runner” (LR) or “Average Joe” (AJ), based on the magnitude and duration of each firm’s performance. Each category’s detailed profile is provided on page 6. A “triplet” consists of three companies from one industry, each with one of these three archetypical performance profiles. We hope first to understand the proximate causes of performance differences within each triplet, and then to generalize across these triplets to identify those aspects of general management that systematically drive significant outperformance.

### Table 1: The Triplet Companies

<table>
<thead>
<tr>
<th>Industry</th>
<th>Archetype</th>
<th>Company</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Appliances</strong></td>
<td>MW</td>
<td>Maytag</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>HMI Industries</td>
</tr>
<tr>
<td></td>
<td>AJ</td>
<td>Whirlpool</td>
</tr>
<tr>
<td><strong>Confectionary</strong></td>
<td>MW</td>
<td>Wrigley</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>Tootsie</td>
</tr>
<tr>
<td></td>
<td>AJ</td>
<td>Rocky Mountain chocolate Factory</td>
</tr>
<tr>
<td><strong>Family Clothing Stores</strong></td>
<td>MW</td>
<td>Abercrombie &amp; Fitch</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>Finish Line</td>
</tr>
<tr>
<td></td>
<td>AJ</td>
<td>Syms Corp</td>
</tr>
<tr>
<td><strong>Machinery</strong></td>
<td>MW</td>
<td>Briggs &amp; Stratton</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>Gorman-Rupp</td>
</tr>
<tr>
<td></td>
<td>AJ</td>
<td>Gardner Denver</td>
</tr>
<tr>
<td><strong>Medical Devices</strong></td>
<td>MW</td>
<td>Medtronic</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>Stryker</td>
</tr>
<tr>
<td></td>
<td>AJ</td>
<td>Invacare</td>
</tr>
<tr>
<td><strong>Motor Transport</strong></td>
<td>MW</td>
<td>Heartland Express</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>Werner Enterprises</td>
</tr>
<tr>
<td></td>
<td>AJ</td>
<td>P.A.M. Transportation</td>
</tr>
<tr>
<td><strong>Pharmaceuticals</strong></td>
<td>MW</td>
<td>Merck</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>Eli Lilly</td>
</tr>
<tr>
<td></td>
<td>AJ</td>
<td>K V Pharmaceutical</td>
</tr>
<tr>
<td><strong>Retail-Grocery</strong></td>
<td>MW</td>
<td>Weis Markets</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>Publix Super Markets</td>
</tr>
<tr>
<td></td>
<td>AJ</td>
<td>Whole Foods Market</td>
</tr>
<tr>
<td><strong>Semiconductors</strong></td>
<td>MW</td>
<td>Linear Technology</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>Micropac Industries</td>
</tr>
<tr>
<td></td>
<td>AJ</td>
<td>Intl Rectifier</td>
</tr>
<tr>
<td><strong>Software</strong></td>
<td>MW</td>
<td>Oracle</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>Symantec</td>
</tr>
<tr>
<td></td>
<td>AJ</td>
<td>Informatica</td>
</tr>
<tr>
<td><strong>Wiring</strong></td>
<td>MW</td>
<td>Thomas &amp; Betts</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>Hubbell</td>
</tr>
<tr>
<td></td>
<td>AJ</td>
<td>Emrise</td>
</tr>
<tr>
<td><strong>Wholesale Trade</strong></td>
<td>MW</td>
<td>Herbalife</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>Aceto Corp</td>
</tr>
<tr>
<td></td>
<td>AJ</td>
<td>AmerisourceBergen Corp</td>
</tr>
</tbody>
</table>
Many others have essayed this question, many using a broadly similar research method, and some have found wide and appreciative audiences for their efforts. But all these efforts — and we say “all” with appropriate circumspection, for it is of course possible that we have missed some — are subject to one or more of the following profound flaws that, we believe, undermine to the point of collapse the confidence one can have in the scientific foundations of their normative claims.

We are hardly the first to question the leading examples of this genre of management research. Phil Rosenzweig’s *The Halo Effect* is perhaps the most piquant criticism of such work. His objections are salutary, and point toward a benchmark of scientific rigor to which one does well merely to aspire.

Our concerns are different from Rosenzweig’s, and the shortcomings we find especially corrosive fall into three broad categories:

**Poorly-selected samples: They’re not examining sufficiently high-performing firms**

The fundamental assumption of many investigations into the defining behaviors of successful companies is that the companies being studied have in fact delivered performance worthy of close inspection. It turns out that this assumption is rarely justified. We have detailed elsewhere the basis of our concern (see *A Random Search for Excellence* at www.deloitte.com/persistence). It is, in our view, perhaps the most damaging criticism one can level against prior work, for if others have been drawing conclusions about excellence based on an examination of the merely lucky, there is little reason to have confidence in the findings.

Perhaps the example that illustrates this in sharpest relief is our analysis of the purportedly “good-to-great” companies in Jim Collins’s blockbuster *Good to Great*. None of those firms has a performance profile that can be distinguished from more than good luck. As a consequence, any utility one might find in his recommendations likely relies on his general business sense rather than his clinical evidence.

**Weak cause-and-effect connections: We are convinced only if we already believe the conclusions**

Examining the behaviors of successful firms and comparing them with those of less successful firms is a credible method of clinical investigation; it is the approach we have adopted. However, many investigators seem to believe that any and all behaviors that are common to firms in a given performance category and unique to that category are therefore relevant determinants of performance. Never (again, as far as we know) has there been a concerted effort to connect the alleged causes of performance with the magnitude of the performance to be explained.

For example, in Joyce et al.’s *What Really Works* the performance differential between “winners” and “losers” is on average a simple return to shareholders of 883 percent over ten years. The researchers claim that the causes of this difference are those characteristics that are most strongly correlated with being a winner. They have taken considerable pains to establish the statistical significance of these correlations, and the features they identified were chosen from a much larger population. There is little in the way of post hoc pattern recognition, and instead bonafide hypothesis testing with valid criteria for determining what was and was not meaningfully correlated with the outcomes of interest.

What is not clear is whether the characteristics so identified are viable causes of a difference in performance of that magnitude. For example, one of their findings is that winners have higher product quality than losers. This seems a reasonable claim, but no attempt is made to show, for example, that higher product quality translates into, say, higher margins, high market share, or higher growth, which in turn generates an incremental return of a specified magnitude over a firm with lower product quality. In other words, although the behaviors in question are correlated with outperformance, and are credible as explanations, there is no quantitative connection between the purported cause and the observed effect. As a consequence, we believe the explanations given because we already think those behaviors are explanations of success.

---

A “black box” analytical method: Trust me!
A defining feature of the scientific method is the ability to replicate results. With the same data and the same method, different researchers should reach the same conclusions. The clinical analyses of excellent companies has been largely resistant to this condition for at least two reasons. First, investigators typically examine vast quantities of data drawn from a wide variety of sources. Many such sources are commonplace — annual reports, major media outlets, investment banking evaluations, and so on. But significant data are often gathered from highly variable sources: interviews, secondary and tertiary media, blogs or websites that come and go, and so on. As a result, the inputs to the analytical process are often unknown, save to the researchers themselves. This makes replicating their analyses effectively impossible.

Second, results are frequently reported simply as results. The analytical process that extracted those results from the underlying data is typically, and often explicitly, each researcher’s own idiosyncratic and often of necessity unspecified “pattern recognition” algorithm. In other words, the answers are whatever they appear to be to whoever is doing the analysis. The implication is that if even if we were able to assemble the same data set, if someone else had analyzed those data we would have gotten different answers — different ultimate causes of success.

Improvement, not Perfection
It is decidedly not our view that the “perfect method” is possible or that, even if it were, the result would be “perfect knowledge.” Our hope is that a different research design - one created with these challenges in mind - will lead to findings that, whatever their substance, will be robust and well-supported.

Central to this design is an opportunity for interested parties to comment on our attempt to explain performance differences among our triplet companies. We will therefore be circulating monographs on our triplets, doing our best to address these and other research design issues.

This prolegomena will serve to describe our overall method and what we hope to accomplish with our triplet monographs. No doubt our work will be found incomplete. Our intent, however, is not to say everything there is to say but rather to have as much confidence as possible in what we ultimately assert.
Not every so-called “success study” is subject to a material degree to each of these failings; but as far as we know, every prior effort is subject in significant enough measure to one or more of them to call into question the scientific validity of their claims. With The Persistence Project we hope to learn from their efforts, designing and executing a research project that is self-consciously aware of the need to address these issues head-on, even as we accept that it is impossible to defeat any of them completely.

Sample Selection
Our full sample selection process is described in A Random Search for Excellence, cited above. In summary, we have specified three categories of performers: Miracle Workers (MW), Long Runners (LR), and Average Joes (AJ).

Miracle Workers: A sufficient number of years in the 9th decile of ROA, conditional on the total number of observations, to be sufficiently statistically unlikely.

Long Runners: A sufficient number of years in the 6th-8th decile band of ROA, conditional on the total number of observations, to be sufficiently statistically unlikely.

Average Joes: Having performance, variability in performance, and lifespan values that are as near as possible to industry median values.

We have been unable to find an entirely objective method for defining superior performance. Consequently, the parameters that define the MW and LR categories are of necessity arbitrary. For example, we could have chosen the 85th percentile or above, rather than the 9th decile. The foundation of our claim of superior sample selection does not, however, turn on this aspect of our study. Rather, it is that firms are said to meet the requirements of a specific profile depending on the “unlikelihood” of having achieved it — that is, their performance is not plausibly attributed to luck alone. Broadening or narrowing the band that defines, say, Miracle Workers would result in a larger or smaller number of firms clearing these benchmarks, which could in turn lead to different conclusions about the causes of “success” — but only because the definition of “success” had been changed. Provided that the test of statistical significance is applied to any given performance profile, the sample would still be credible.

These three categories have been defined to provide contrasts among different performance profiles of interest: the very best, the very good, and the entirely average. Comparing each with the other will, we believe, allow us to determine what separates each from the other. Those evaluating our research must decide for themselves whether the hurdles we have defined for each category are a credible foundation for meaningful results.

Cause and effect
Within each triplet we look for behaviors that differentiate the MW, LR, and AJ from each other on the premise that only differences in behavior can explain differences in performance. This is consistent with the research design of other studies in this genre.

Perhaps uniquely, however, we will attempt to connect the magnitude of the effect of these behaviors with the magnitude of the performance differential to be explained. For example, should we observe that “developing international markets” is a difference in behavior between two firms, we will attempt to assess the impact of incremental penetration of international markets on ROA, and compare the observed performance difference between the LR and MW over the relevant period.

This does not mean that it was definitively the incremental internationalization that caused the performance difference. Peering too closely at what it means for one thing to cause another rapidly causes metaphorical eyestrain — it is a vexing issue that has stymied philosophers from David Hume onward. Our claims of causality will necessarily be circumspect, and their credibility depends on each reader’s subjective assessment of the following two elements:
Temporal relationship: The cause must precede the effect by a relevant period. For example, if a MW is observed to execute more mergers than a LR, then for that difference in behavior to be a credible cause of differences in performance the incremental performance to be explained must follow the mergers by some meaningful period of time. Rarely, for example, would a merger in one year have material positive impact on performance the following year. And if that is the claim to be made, special care must be taken to explain how such an impact was observed so quickly.

Relevant and plausible impact: It is unnecessary to approach research into the causes of superior performance with a blank slate. Not all observed differences in behavior between, say, a MW and the relevant LR are plausible causes of differences in performance; there will simply be too many behavioral differences, and our outcomes will be grossly over-determined. Thankfully we can draw on the great deal that we already know about how specific behaviors affect performance generally.

For example, we know that differentiated product market positions have a better chance of generating higher margins than cost-based advantages. We know that there can be significant lags between investments in research and development and the launch of new, successful products. We know that mergers and acquisitions have a high variance in outcomes, and that larger deals are especially fraught. Consequently, explaining a large margin advantage in terms of rapid cost synergies extracted from a huge merger requires much stronger anecdotal evidence than observing instead the final fruition of a years-long R&D initiative that led to the launch of a raft of new and highly differentiated products.

In other words, our explanations must not only be able to account for the performance observed, but must square with what is already known about how the world works.

Replicable and transparent analytical method in our monographs

This last element is what each triplet monograph is most in service of, for we present our explanation of the differences in performance among our MW, LR and AJ in a given industry.

In an attempt to make our analytical method more transparent, and improve our results, we are circulating for comment our triplet-specific explanations of the observed performance differences, soliciting input from a wide range of interested parties: investment analysts, industry insiders, company experts… and you. Working across a dozen industries necessarily means that we are less than expert in all of them. Despite dozens of person-years of effort, our explanations will undoubtedly be incomplete. And if we were to attempt cross-case analysis on those incomplete explanations, these inevitable shortcomings would merely be compounded. We have no doubt that by soliciting input at this early stage of our analysis we will only improve the validity of our triplet-level explanations. Our ultimate goal is that the development of our final conclusions will rest on a tested and validated set of observations within each industry.
To facilitate the ultimate generalization of findings within each triplet, we have adopted a consistent analytical approach to understanding performance differences and their causes.

Objective
For each triplet, we are seeking explanations that account as completely as possible for the observed performance differences. These explanations are created with no consideration to their generalizability across companies or industries.

Each triplet analysis, then, should be seen as its own “data point” in the larger project. Our intent is to describe proximate causes of superior, persistent performance for each triplet, and then generalize across the sample of triplets with the hope of identifying more general characteristics of strategy, operations, organization, management, decision-making, or other elements that emerge from a cross-case comparison.

Approach
Although we characterize a firm as a MW, LR, or AJ in virtue of its lifetime performance (where “lifetime” means all available data from 1966–2006) we are not looking for behaviors that are consistent across that entire period to explain overall performance. Across a forty-year period there are inevitably periods of higher and lower absolute and relative performance and each will require an explanation.

We have adopted, and adapted for our purposes, a “structure-conduct-performance” analytical framework. Following generally accepted practice, we go beyond the classical formulation of the “structure-conduct-performance” paradigm: there is more to industry structure than merely the number and relative size of firms; there is more to firm behavior than merely pricing and output within a given product market; and there is more to firm performance than simply its ROA.

Our analysis would have a certain theoretical elegance if we were able to move linearly through these three elements. A description of industry structure would be followed by an identification of relevant behavioral differences, which would then give rise to an assessment of the nature and magnitude of the financial impact of those differences.

Unfortunately, the world is too complex a place to adopt such an approach. There are simply too many possible candidates for behavioral differences. Was it the structure of the board of directors? The approach to M&A? The CEO succession planning? The level of diversification? The organization structure? The surfeit of potential explanations is, in our view, a large part of the reason so few “success study” efforts result in similar findings — which makes having confidence in any given success study very difficult.

To try and mitigate this problem we have adopted the view that the plausibility of an observed difference in behavior as an explanation for an observed difference in performance is constrained by two sets of considerations: the financial structure of the performance differences, and the industry context in which the performance difference was realized. The first of these is the more constraining, and so in order to avoid chasing too many red herrings, we begin our triplet analysis first with a careful investigation into the nature of the performance difference itself: is the return on assets advantage driven by margin or cost advantages, and within each of these, what revenue and cost elements drive them during a particular period? This directs us toward those activities that are likeliest to be consistent with particular financial outcomes.

Then we examine industry structure, and how it has changed over time. High level industry trends, such as market-shaping regulation, new technologies, or consolidation often have material influence on what sorts of behaviors are likely to drive superior performance. Consequently, examining differences in behavior that are affected by these forces is likely to be an especially fruitful line of inquiry.
Guided by this deeper understanding of the structure of performance differences and competitive context within which these differences were realized, we then identify potentially relevant performance differences and assess their materiality. This identification stage rests as much as possible on hard data, and our assessment of materiality is based on financial modeling that asks the question “what if our MW had behaved as did our LR/AJ?” If our estimates of the financial impact of the behaviors we have identified suggest that we have explained a material portion of the observed performance difference in a manner that is consistent with both the structure of the financial performance and the larger competitive context, we will be willing to conclude that we have a behavior that matters.

Consider now each of performance, structure, and conduct in greater detail.

Performance
We begin by identifying performance “eras” based on the relative differences in performance among our three firms and trends in those differences over time. When is our MW doing particularly well? Are there periods when the MW’s advantage is waning or waxing? Does our LR ever enjoy a period of particularly noteworthy performance? When is the AJ’s performance especially volatile? The eras provide some guidance on the timing and magnitude of performance impact that differences in behavior must have in order to be useful explanations. This guides our clinical research so that we avoid focusing on every difference we can find, and instead concentrate on those differences that, individually or collectively, have a chance of being important enough.

The cumulative ROA differences among our three firms in a given era are decomposed into their constituent elements, return on sales (ROS) and total asset turnover (TAT):

\[
\text{ROA} = \frac{\text{Income}}{\text{Assets}} = \text{ROS} \times \text{TAT}
\]

Each of ROS and TAT is then be further decomposed into its major constituent elements:

**Return on Sales**
- Cost of goods sold (COGS)
- Sales, general & administrative (SG&A)
- Research and development (R&D)
- Taxes
- Non-operating income
- …and so on.

**Total Asset Turnover**
- Fixed asset turnover
- Current asset turnover

This additional detail provides clues to further guide our clinical work. For example, was the lion’s share of the MW’s superior ROA a function of better ROS? Was that ROS driven by lower COGS? Was higher ROS maintained even with higher R&D spending? Was the AJ suffering from sustained and much lower asset turnover, making it difficult to drive ROA despite lower SG&A?

There are, of course, many different ways a company might realize any given combination of these profit and cost elements to realize its overall ROA. For example, a low COGS percentage that contributes to a high ROS can be a function of either high margin or low costs. Similarly, a high asset turnover ratio can be a consequence of strong sales growth or particularly efficient asset management.

Even so, the overall pattern of the composition of performance over time does narrow down considerably the range of possible explanations, and provides a valuable benchmark for assessing the consistency of our clinical findings with the hard facts of financial performance.

Structure
The plausibility of any particular explanation is further constrained by its consistency with important features of the larger competitive context — the industry structure. Our review of each industry is intended to capture the salient features that have the most credible claim on affecting the fortunes of the industry’s major players. These include, but are not limited to, changes in regulatory regimes, the rise of new competitors, technological innovation, and the globalization of an industry. We have attempted to provide compelling evidence in support of each putatively significant change.
Constraints on the plausibility of an explanation arising from considerations of industry structure might be changes in distribution channels as customers consolidate, resulting in increased customer bargaining power. A significant shift in this feature of the competitive landscape is a good candidate for a behavioral source of performance differences. Of course, there is no guarantee that our MW, LR, and AJ will have actually pursued different responses, but an awareness of such structural considerations provides a useful backdrop and starting point. Alternatively, should a MW or LR deliver its exceptional performance seemingly in spite of behaviors that seem at odds with structural considerations, we will know to look more closely at the full repertoire of behavioral responses in order to identify what other actions were able to compensate for the seeming misalignment.

**Conduct**

Our investigations are into the causes of differences in performance among the firms in our triplets. Consequently, our focus when describing the relevant conduct of the firms in a triplet are the differences in behavior among them.

In establishing these differences, the primary basis of our behavioral claims will be observable and, where possible, quantifiable facts. Subjective assessments of a company’s behavior can be materially polluted by “halos.” A company’s patent output is a fact, and supports claims of productive R&D; being named to BusinessWeek’s list of “Most Innovative Companies” is not. (Being so-named might be a clue that it is worth exploring whether a company is in fact innovative, but it is not evidence that it is innovative.)

We pay a high price for this restriction. In the same way that paleontologists can conjecture about, say, what sorts of noises dinosaurs made or what their social structures might have been, they must restrict the claims they make with confidence to what they can justify based on the fossil record. And since, generally, only some of the hard bits fossilize, dinosaur hunters work with a small and biased sample. Similarly, we will be working with only a small fraction of all the data available, and those data we have will be biased strongly toward the quantifiable and more-nearly objective.

Perhaps the most significant delimitation this restriction in method places on the significance of our findings is how much of the observed variance in performance we will be able to explain. In some years, for some comparisons, we are able to explain a large part of this difference. In other years we explain a relatively small part of the difference. When we explain most of the difference, and we have met the restrictions of imposed by our decomposition of financial performance and our identification of key industry trends, we can be confident that our explanation is important. When we explain comparatively little, we must conclude that the less objectively observable elements of behavior are driving performance, and that what is truly distinguishing one firm’s performance from another remains unknown. But such are the wages of aspiring to scientific credibility.
Concluding thoughts

The plausibility of our explanations will turn on the degree to which they fit with the key factors driving success in the industry at that time and whether they can account for the magnitude and nature of the performance differences observed. There can be no dispositive proof for this or that particular explanation: the world is too multi-faceted and causal linkages are too ambiguous. Our explanations will turn ultimately on the cumulative weight of the circumstantial evidence. But we believe that by explicitly connecting our explanations to both the industry-level determinants of success and the company-level performance to be explained we can build a convincing enough clinical case to support the claims that our identified behaviors are at least significant contributors to significant performance.

Central to this project is creating credible, accurate explanations of performance differences among our triplets. This prolegomena has described how we hope to achieve this. As these monographs are published over the coming months, we invite you to join us in that quest.
Authors

Michael E. Raynor
Deloitte Consulting LLP
mraynor@deloitte.com

Mumtaz Ahmed
Deloitte Consulting LLP
muahmed@deloitte.com

About Deloitte

Deloitte refers to one or more of Deloitte Touche Tohmatsu, a Swiss Verein, and its network of member firms, each of which is a legally separate and independent entity. Please see www.deloitte.com/about for a detailed description of the legal structure of Deloitte Touche Tohmatsu and its member firms. Please see www.deloitte.com/us/about for a detailed description of the legal structure of Deloitte LLP and its subsidiaries.

Copyright © 2010 Deloitte Development LLC. All rights reserved.
Member of Deloitte Touche Tohmatsu