1 The Informational Advantage

Timely, pertinent and accurate information is critical to effective decision making. Conceptually, this claim is intuitively obvious to most of us, yet, having access to “right” information is hardly the norm in many business decision contexts. The reason for that is that as much as the idea of timely and accurate information being readily available is compelling, making it a reality is an involved and a potentially complex undertaking. Still, even the most operationally challenging problems can be solved, as evidenced by a number of organizations that were able to develop superior decision-aiding informational resources.

This chapter tackles the key conceptual and operational challenges of developing superior informational resources. It begins by expressly differentiating between “information” and “knowledge,” with the goal of drawing attention to a common—and incorrect—practice of using these two terms interchangeably, followed by the discussion of “knowledge imperative,” which captures the ever-growing importance of decision-aiding knowledge to business decision making. Lastly, the emerging field of database analytics is also discussed.

The Age of Data

Knowledge is the currency of the Information Age. In business, it is the foundation of effective decision making, and as such, an increasingly dominant predictor of organizations’ success. Across industry sectors, firms that “know more” earn higher returns on their invested capital, simply because they can focus their energies and resources on the most beneficial courses of action. Within the realm of product and service promotions, superior insights enable organizations to grow the size and profitability of their customer bases by more effective targeting and fine-tuning of individual promotional mix vehicles. In short, better customer and promotional mix knowledge leads to greater productivity of the organization's promotional efforts.

Much of the decision-aiding knowledge comes from insights derived from the so-called “transactional” data, which encompasses a wide range of business-to-consumers and business-to-business interactions.[1] Thanks in a large part to the brisk growth of *electronic transaction processing* (ETP) systems, transactional data became ubiquitous over the past couple of decades. Starting with the introduction and subsequent proliferation of barcode-reading electronic scanners (first used in an Ohio supermarket in 1974), and further fueled by rapid advances in computing and data storage technologies, the ETP systems now generate enormous amounts of relatively accurate and detailed business transaction-describing data.

ETP was taken to a new level by the explosive growth of online commerce, which contributed volumes of transactional and related (such as the so-called “click-stream”) data not seen before. Online commerce, however, is not just yet another source of transactional data—it also enables the levels of “trackability” of promotional imprints that were previously not possible. In other words, in contrast to the (very) hard to estimate number of impressions associated with television or billboard advertising, online promotions afford a far greater precision of promotional exposure estimation. Ironically, the otherwise virtual world of online business can be very tangible when it comes to the basic "who," "what" and "when" of business interactions (which spurred a rather intensive information privacy debate).

Another important aspect of the widespread adoption of ETP as the primary mode of processing business transactions has been the commoditization of the resultant data. Stated differently, most organizations competing in a given industry have comparable data assets, hence it follows that any cross-organization data-related knowledge differences are most likely a function of data analytical advantages. Just about every organization has lots of data, but some are simply better at squeezing competitively advantageous insights out of it.

It seems fair to say that, at an organization level, data crunching capabilities evolve primarily in response to organizational priorities and the growth in the availability of data. It follows that as more data becomes available, data crunching capabilities of organizations will also steadily expand. However, since the expansion of data analytic capabilities is also contingent on organizational priorities, not all companies will develop at the same pace. In the end, a common pool of data brought about by the aforementioned proliferation of electronic transaction processing systems (such as barcode readers widely used in retail) will bring about some common-to-all, generic data analytical capabilities, while at the same time creating competitive disparity in terms of more advanced data analytical competencies.

This trend is particularly evident in the area of marketing promotion evaluation. The vast majority of organizations that have access to the requisite data will typically also have "basic" data analytical capabilities, usually built around data summarization and reporting. Chances are that marketing managers at firms competing in the same industry are looking at very similar sales reports. Much of that basic informational parity can be attributed to the proliferation of third-party developed reporting applications, such as the widely used "business intelligence" tools offered by a number of different vendors, such as Business Objects, Cognos or MicroStrategy.

By all accounts, the convergence of widespread data availability and reporting capabilities should have produced a leveled informational playing field—in other words, most firms competing in a particular area ought to have comparable informational competencies. Yet, that is not the case. As detailed below, in virtually all industries only a handful of firms are able to consistently use data—readily available to most—to their advantage. To paraphrase an old cliché: *Most companies are* (still) *data-rich, but information-poor*. Even though data is accessible to the vast majority of organizations, it tends to widen the competitive divide rather than narrowing it. Hence, as pointed out in a growing number of popular business texts, such as Davenport's *Competing on Analytics* (2006), Levitt and Dubner's *Freakonomics* (2005) or Ayres' *Super Crunchers* (2007), advanced analytical "know-how" has become one of the key determinants of firms' marketplace competitiveness.

As noted earlier, in the knowledge-intensive environment, informationally competent firms are able to consistently outperform their competitors, primarily because they are able to make better use of organizational resources. Whether it is better understanding of consumers' needs and preferences, a more accurate assessment of the impact of competitive actions or more impactful allocation of promotional dollars, better information typically leads to better decisions. Knowing less, on the other hand, tends to introduce an element of randomness into the organization's decision making process (as the shortage of robust information necessitates guessing), which over the long run translates into a more uneven organizational performance. And last but not least, informational competency enables organizations to take a more proactive decision making stance, which is generally viewed as a prerequisite to both winning and maintaining market leadership.

The lack of reliable decision insights tends to impose a more reactive decision making mode, which in turn tends to force organizations into playing catch-up with their better informed rivals.

It is important to note that persistent informational deficiency does not just negatively impact the organization's performance—it may actually pose a threat to its very survival. The steadily accelerating pace of globalization coupled with the broadening trend of deregulation continues to stiffen the competitiveness of markets, which in effect is raising the cost of poor decisions. Under most circumstances, there is simply too little time and too much competition to practice "trial and error" decision making.

As demonstrated by the likes of Microsoft, Proctor & Gamble or Marriott, timely, accurate and unique business insights are among the key pathways to sustainable competitive advantage. The degree to which market leaders are able to consistently outperform their competitors is now inextricably tied to their proficiency in translating large volumes of data into decision-guiding insights. The speed and precision with which an organization can translate raw data into decision-guiding insights determines whether it will be able to pinpoint competitive weaknesses and to identify the most advantageous courses of action, or be among the first to spot and take advantage of emerging trends and opportunities. And as the marketplace competition continues to heat up, fueled by growing privatization, accelerating product innovations and the widening trend of deregulation, it is not just the success, but even the very survival of organizations that is becoming increasingly more dependent on their ability to make sound and effective decisions. In a sense, all organizations are now in the information business but their competencies in that area are highly uneven.

In view of the enormous scope and the depth of what can be included under the broad label of "marketing analytics," this book does not pretend to offer a "one size fits all" solution that could be applied to all data-analysis-related business problems. In fact, given the wide range of potential topical areas, industry- and company-specific circumstances and the types of data, such a "theory of everything" solution does not appear feasible, at least at this time. With that in mind, this text is concerned with one particular area of business knowledge—the "how-to" of efficient, objective and reliable promotional programs evaluation. The ideas put forth here are built around the belief that organizations can use rational planning and evaluation processes and analytic techniques to optimize the productivity of their promotional mix. More specifically, thoughtful promotional managers can increase the net revenue contribution of their promotional mix. The journey toward that end objective starts with an explicit investigation into what constitutes knowledge.

The Believability Factor

There are some very hard to believe facts associated with folding an ordinary sheet of notebook paper. First, regardless of the size of a sheet, no one has been able to fold a sheet of paper more than twelve times.[2] However, what is even more extraordinary about paper folding is the height of the resultant stack. Starting with an appropriately sized sheet of ordinary notebook paper, folding it seven times (the number of folds once believed to constitute the upper limit) will result in a stack approximately equal in height to the thickness of an average notebook. Another three folds will result in the stack height about the width of a hand (thumb included), and an additional four (for a total of fourteen folds) would push the height of our stack to be roughly that of an average person. If we were to continue to fold our sheet of paper, the expected results become very hard to believe: Seventeen folds would produce a stack the height of an average two storey house; another three folds (for a total of twenty) would yield a stack reaching approximately a quarter of the way up the Sears Tower. If folded over thirty times, the resultant stack would reach past the outer limits of Earth's atmosphere; and lastly, if folded fifty times, our ordinarily thin, albeit extraordinarily large in terms of area (to allow a large number of folds) sheet would produce a stack of paper reaching … all the way to the Sun. That is roughly 94 million miles!

For most of us, years of schooling imprinted our minds with a variety of abstract notions, while also conditioning our psyche to accept a considerable amount of intangible truths. So long as those scientific and other truths do not come in conflict with our "common sense" of reality, most of us are generally willing to accept even somewhat far-fetched claims. However, when that is not the case—that is, when a particular claim violates what we consider to be reasonable, the result is *cognitive dissonance*. We just can't accept a particular fact or a claim as being true. Even if the underlying rationale and the empirical method both seem acceptable and correct, it can be still very, very hard to believe a conclusion that "does not make sense". That is precisely the case with the paper folding exercise. It is an example of *exponential growth*, which is a phenomenon where the rate of growth rapidly increases as the quantity (e.g., the above stack of paper) gets larger. Since it is a well-defined mathematical property we can compute its values without the need for physical measurements, which is the reason we are able to estimate the height of the stack of paper, even though we are not physically able to fold a sheet of paper fifty times. I am going to venture to say that those of us who at some point in our educational journey were exposed to the notion of exponential growth found it to be intuitively clear and reasonable; furthermore, once properly explained, the computational steps also made sense, which is to say their logic did not clash with our view of the world. Yet when put to a bit of an extreme test, that otherwise acceptable concept can yield unacceptable conclusions. Folding a thin sheet of paper a relatively small number of times simply cannot result in such a staggeringly high stack…

This example underscores both the value and the challenge associated with using data analysis derived knowledge as the basis of decision making. It is very easy to accept the findings that fall in line with our existing beliefs, though it could be argued that little incremental value comes out of such "discoveries". It is altogether a different story when findings contradict our a priori beliefs: Is there a problem with the data? Is the approach flawed? Are there any errors…? To be fair, data can be corrupted, an approach can be flawed and we all certainly make mistakes. At the same time, however, if none of that could be the case—what then? Oftentimes, doubts linger and what could have become an inspiration for a competitively advantageous decision, joins the repository of many other research initiatives, all dutifully written up, but never acted upon.

Yet taking the leap of faith and acting in accordance with objectively validated insights can be quite beneficial. Much of information technology that permeates our professional and personal lives is "powered" by quantum mechanical predictions; in fact, quantum theory is, in terms of the accuracy of its predictions, the most accurate scientific theory ever constructed.[3] At the same time, it is among the most bizarre, hard-to-believe frameworks in terms of its postulates. In the quantum theoretical world, objects can exist in two states or places simultaneously (a condition known as "superposition"), in addition to which, objects are also instantaneously "aware" of distant other objects (an effect known as "quantum teleportation"). It is akin to saying that a person can be simultaneously alive and dead and furthermore, that a person's physical existence is entangled with consciousness of others. What then determines whether someone is alive or dead? The act of looking, stipulates quantum mechanics. In other words, perception creates physical reality. Does that sound believable?

To Einstein these were "spooky interactions" which is a term he coined deriding the quantum theory. In fact, the great scientist spent more time trying to disprove the quantum theory than he did crafting his own theories of general and special relativity. But in the end, he failed. As much as it is hard to intuitively come to terms with the bizarre postulates of the quantum world, the equations describing its mechanics are extremely reliable. Microchip-powered computing devices, like the laptop on which I am typing this text, work undeniably well, because of the accuracy of quantum mechanical predictions, even though most of us have very hard time accepting the picture painted by the underlying theory.

Obviously, this is not a text on quantum mechanics or paper folding trivia. However, these two examples point to an interesting assertion: The believability of analytically derived explanations should not always be the ultimate determinant of whether or not we accept—and more importantly, act upon—the findings. This is not to say that we should totally disregard our intuition, for that would mean depriving ourselves of lifetime worth of accumulated, though not always well catalogued, knowledge. Quite to the contrary—I am arguing that true edge-producing knowledge needs to combine the elements of truths that might be intuitively obvious to us with those that may not make sense to us, but have been shown to be empirically true. In other words, why not try to get the best of both worlds?

Data, Knowledge and Decisions

As ably detailed by Quinn and his colleagues,[4] success of organizations depends more on their intellectual and systems capabilities than physical assets. To a large degree, Quinn's conclusion is somewhat intuitively obvious: Physical assets are, for the most part, generic, thus it is the application or deployment of those assets that determines the overall success of organizations. Stated differently, it is the uniqueness of organizational "know-how," coupled with the ability to utilize that knowledge that are the greatest influencers of success. Hence it is of considerable importance to organizations to systematically develop competitively advantageous insights in a way that will aid their decision making processes.

It is not an easy task. Unlike the objectively measurable physical assets, knowledge is highly abstract and difficult to measure, both in terms of quality as well as quantity. In the organizational context, it is often either confounded with individuals or buried deep inside various reservoirs holding it. An even more fundamental challenge is knowing what we know, especially given the pervasiveness of the use of terms such as "data," "information," "facts," "insights" or "knowledge." When is what we know an inconsequential (as far as the ability to enhance the quality of decisions), informational tidbit and when is it a true, difference making insight? The next section hopes to provide some clarification.

What is Knowledge?

Plato[5] defined *knowledge* as "justified true belief." The *Oxford English Dictionary* lists several different definitions of knowledge: 1. "expertise, and skills acquired by a person through experience or education; the theoretical or practical understanding of a subject,"2. "what is known in a particular field or in total; facts and information," and 3. "awareness or familiarity gained by experience of a fact or situation." Wikipedia offers probably the simplest definition, by equating knowledge with "what is known."

Oddly (in view of its 2,500 years or so vintage), Plato's definition of knowledge comes the closest to what it means to business decision making. In business in general, and marketing management in particular, decision making is necessitated by a plurality of alternatives—if there are no alternatives, defined here as substitutable courses of action, there are no decisions to be made. In view of that, knowledge can be construed as the degree of understanding (of relative advantages and disadvantages) of the competing options, or in Plato's terms, beliefs that are “justified and true” regarding the value of alternatives. The possession of such robust understanding leads to selecting the “best” alternative, or the option yielding the greatest net benefit.[6] Hence from the standpoint of marketing decisions , *knowledge represents justified and true beliefs regarding the relative efficacy of competing courses of action*.

Taking this line of reasoning a step further, knowing more will give rise to *informational advantage*, which is *the ability to make more effective decisions stemming from better understanding of the potential outcomes of competing courses of action*. In other words, deeper insights or better know-how on the part of promotional management will contribute to the organization's competitiveness—in fact, under some circumstances, it could be a source of competitive advantage. However, in order to have that type of a profound impact, organizational knowledge has to exhibit several broadly defined characteristics—most notably, it needs to be codifiable, teachable and systemic.

Codifiability of knowledge is its ability to objectively encode facts and inferences into clear behavioral guides. In a practical sense, it is a degree to which a particular set of insights or know-how can exist independently of those creating it. Examples of codifiable knowledge include multivariate statistical models-generated behavioral expectancies or propensities, such as the probability of adverse development of recently filed liability claims or the likelihood of securities or employment practices class action. On the other hand, marketing managers' experience or intuition are not easily, if at all, codifiable. As discussed in the next section, not all knowledge can be encoded and as a result, communicated.

The *teachability* of knowledge reflects the degree to which it can be absorbed by the organization. In general, the more understandable and parsimonious the knowledge, the simpler it is to teach. That said, we sometimes do not draw a sufficiently clear line of demarcation between knowledge creation and its application, which is particularly the case in business analytics, where many potentially valuable insights are lost in the web of methodological complexities. In most circumstances, there is a significant difference between teaching how to conduct analyses and how to use the results, which is an important distinction explored in more detail in subsequent chapters. For now, let it suffice to say that insights communicated in a user-friendly format and dispensed in manageable quantities tend to be easy to absorb by the organization, which means that over the long haul they will have more impact on decisions.

Lastly, knowledge has to be *systemic*, which is to say that it needs to permeate the organization. This is important simply because organizations are effectively systems of diverse functions that need to act in harmony to meet its stated objectives. For example, in order for the knowledge of the expected impact of marketing activities to contribute systematically to the growth in revenue and/or profitability, it needs to be made accessible (and taught) to multiple organizational functions, such as marketing and brand management, CRM (customer relationship management) and others.

Components of Knowledge

The creation of knowledge will continue to be a largely human endeavor for the foreseeable future. I am not trying to say that information technology will not play a progressive bigger role in the development of organizational know-how; instead, I am trying to draw attention to an important distinction between *knowledge* and *information*, which are often used interchangeably (both notions are discussed in more detail in the *Data Basics* chapter). Definitionally, information is best described as facts placed in a context, while knowledge is the sum of interpreted information—in other words, information constitutes input while knowledge is the final outcome. It means information is singular and for the most part non-evaluative (i.e., it contains no cause–effect delineation), while knowledge is cumulative and interpretive (i.e., observed outcomes are presented as results of specific actions). Hence, while information technology will certainly play an ever-increasing role in the generation and dissemination of information, the creation of decision-guiding, competitively advantageous knowledge is simply too complex, and to some degree too intangible to be automated, at least in the foreseeable future.

The cumulative character of knowledge gives rise to what is often called *explicit knowledge*, which is factual, objective and relatively easily codified and communicated. In essence, it is an analog to a database. The interpretive dimension, on the other hand, is evident in what is known as *tacit knowledge*. It is a subjective though uniquely human aspect of knowing that is hard to codify, teach and systematize, exemplified by skills, intuition and experience (hence the aforementioned intangibility of knowledge creation). In short, what we tend to regard as knowledge is essentially a product of the interplay between "hard facts" and "fungible interpretation," as shown below in Figure 1.1. Combining these two quite different though equally important dimensions is the primary difficulty in automating the creation of knowledge processes, which is the reason for my earlier claim that the creation of knowledge will remain a largely human endeavor, at least in the foreseeable future.

A somewhat different way of thinking about the building blocks of knowledge is to look at it from the standpoint of epistemology, which is a branch of philosophy concerned with the nature and scope of knowing. An epistemological definition points to four distinct types of knowledge: logical, semantic, systemic or empirical. *Logical knowledge* is the result of the understanding of how ideas relate to one another. Here, knowing manifests itself in applying accepted rules of logic, which stipulate additional truths. For example: All human beings are fallible. John is a human being, therefore, John is fallible. *Semantic knowledge* is the result of learning the meaning of words, which is simply the familiarity with definitions. The definition of epistemology mentioned earlier is an example of semantic knowledge. *Systemic knowledge* is the learning of a particular system, its symbols and their interpretations. For instance, one's mathematical skills are an example of systemic knowledge. Lastly, *empirical knowledge* is the learning resulting from our senses. Much of the scientific knowledge falls under that umbrella—in the fact the scientific method discussed later relies heavily on empirically derived understanding.

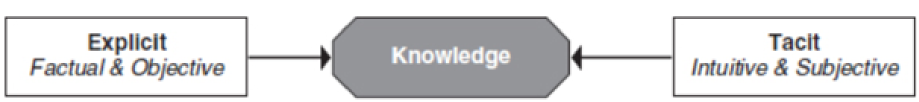

Figure 1.1 The Components of Knowledge

Yet another way of categorizing knowledge is to think of it in terms of a broadly defined purpose it serves, which can be either descriptive (also called declarative) vs. procedural. The former captures the essence of our understanding of "what is," while the latter encompasses our understanding of "how to" accomplish something. For example, knowing the frequency of certain types of loss-generating events and/or the severity of those events constitutes descriptive knowledge; knowing how to reduce the said frequency and/or severity exemplifies procedural knowledge. In the area of marketing database analytics, much of what falls under the umbrella of descriptive knowledge constitutes relatively generic information—the truly competitively advantageous insights usually exemplify procedural knowledge.

Knowledge Creation

How is knowledge created? Although specific mechanisms are probably too diverse and numerous to summarize here, knowledge creation can be either a conscious, end-objective-guided endeavor or it can be a result of an unintended "accident." In other words, knowledge is created purposefully or incidentally. The former is quite familiar to most in the business world, as it is the primary means of generating decision-guiding insights in business. For instance, we notice a sudden upturn in the sales of a particular brand, which then precipitates the question of "why." The process of answering this question is essentially the process of creating purposeful knowledge. On the other hand, the insight that ultimately led to the invention of the now-ubiquitous microwave oven was an unintended by-product of unrelated research into physical properties of very short wavelengths, i.e., microwaves.[7] The researchers were not seeking to gain an understanding of the heating or cooking properties (they were researching radar properties of microwaves; in fact, the first commercially sold microwave oven was called Radar Range), and it was a pure accident that one of the researchers put a candy bar in his shirt pocket which began to melt as a result of direct exposure to microwave radiation…

Stories like the accidental microwave invention certainly stir our imagination, but—at least in business research—it is the often painstaking, systematic, purposeful pursuit of knowledge that can contribute to the creation and sustainment of competitive advantage. Stated differently, to be effective, analyses of business data should be directed toward answering specific questions. Thus as approached in this book, knowledge creation is a teleological[8] process, which needs to be directed at a specific purpose to yield worthwhile results.

It all seems straightforward, though it is not. Most organizations are far more adept at capturing data than they are at turning it into decision-aiding insights; in fact, the sheer volume and diversity of the available data can get in the way of using its informational content productively. Strategies often get lost in a sea of "interesting," albeit accidental findings generated and disseminated not because they support specific objectives, but simply because they are…well, interesting. More importantly, these often trivial informational pursuits tend to draw the same level of vigor (frankly, oftentimes even more as "interesting" tends to be more captivating than "important") as does the pursuit of insights to guide the firm's stated strategic goals. Hence in some instances it is not the scarcity but the overabundance of data that impedes the creation of competitively advantageous knowledge, which when coupled with ineffective information filtering processes can significantly diminish the potential value of corporate data.

And thus the challenge: To create a hard to imitate knowledge base to serve as a foundation of sustainable competitive advantage, by means of injecting unique insights into the organization's decision making processes. As pointed out earlier, it means the pulling together of the two, frankly quite different dimensions of knowledge: explicit and tacit.

Starting with the former, the creation of a robust *explicit knowledge* reservoir requires the translation of data-derived information into higher level insights. This entails two somewhat different and temporally sequential steps. First is the *informational reduction*, which is a set of statistical procedures-based activities (discussed in more detail in subsequent chapters), designed to expressly differentiate between facts that are critical to reaching the stated strategic objectives and those that are not. The second step is that of *meta analysis*, which entails summarizing the critical results (based on informational reduction), but still too granular information into higher-order insights. The entire data-to-explicit-knowledge process is depicted in Figure 1.2.

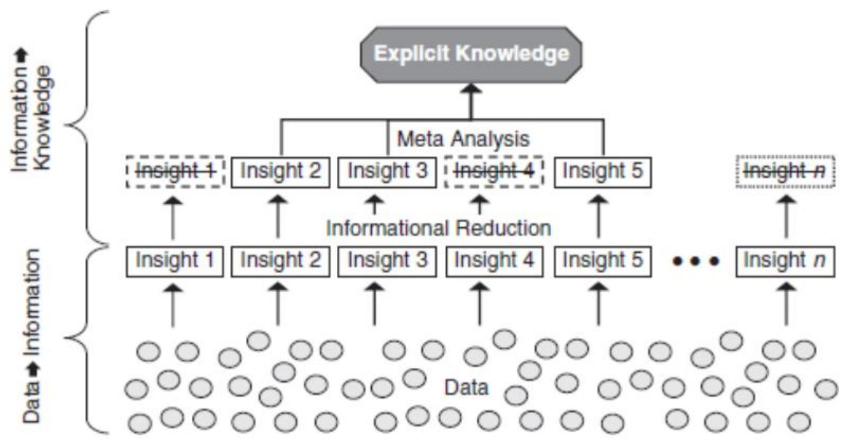

Figure 1.2 Creating Explicit Knowledge

Explicit knowledge alone, however, rarely gives rise to competitive advantage because it is still devoid of experience-based fact interpretation, thus it is usually not sufficiently indicative of the most advantageous courses of action. In other words, the absence of the experiential dimension, such as an experience in a particular industry to help contextualize company-specific loss experience, or hands-on claim management or loss-prevention experience may both limit the potential applicability of the fact-only-based explicit knowledge. This is particularly important considering that in many industries most of the competitors have access to a wide array of fundamentally the same information; hence, working from the same factual base, they are likely to arrive at similar insights, if their knowledge creation pursuits are limited to the explicit dimension of knowledge.

Adding the experiential or tacit component of knowledge into the knowledge creation mix helps to draw attention to what matters from a competitive standpoint. Thus to gain informational advantage over its competitors, an organization has to develop proprietary sources of factual information or find an effective method of “personalizing” its factual knowledge by combining it with the elements of *tacit knowledge*. There are numerous examples of successfully pursuing both alternatives. The world's largest retailer, Walmart, in contrast to most of its competitors has consistently refused to sell its store movement data (i.e., sales data collected from the point-of-sale devices) to the outside syndicated information aggregators (most notably, AC Nielsen and Information Resources) to assure the uniqueness of its knowledge. The results are self-evident… On the other hand, Capital One Bank, working fundamentally with the same type of data as its competitors invested heavily in its own, proprietary data analytical processes (i.e., tacit knowledge) which enabled it to consistently deliver above average financial results. The competitive and other circumstances in which these two companies operate pushed them in somewhat different directions, but both organizations made the most of their circumstances, largely because they were purposeful and systematic.

Harnessing the value of tacit knowledge is in many regards more challenging as so much of it is subjective and difficult to codify. That said, a systematic process, framed in the larger context of marketing analytics can serve as a conduit to extracting, normalizing and ultimately, incorporating the tacit dimension into the overall organizational knowledge base. The recommended process is depicted in Figure 1.3.

As suggested earlier and graphically shown in Figure 1.3, the broadly defined analytical process can be construed as a conduit to systematically transforming data into information and, ultimately, (explicit) knowledge. Furthermore, it is also as a source of the more experience-based tacit knowledge, which is important from the standpoint of maintaining the objectivity of the resultant insights. In other words, the data analytical skill set (*analytical process*), coupled with accumulated industry and related expertise (*experience*) support both the process of extracting explicit knowledge out of data (*data* → *information* → *explicit knowledge* progression) and the accumulation of tacit knowledge. The end goal of the knowledge creation efforts—which is sustainable *competitive advantage*—is the product of a unique combination of explicit and tacit knowledge at the organization level. In a conceptual sense, the entire process outlined above is a reflection of the teleological nature of business knowledge creation and accumulation: competitively useful knowledge creation is directed at a specific purpose, rather than "fishing" for whatever can be found in the seemingly endless ocean of data.

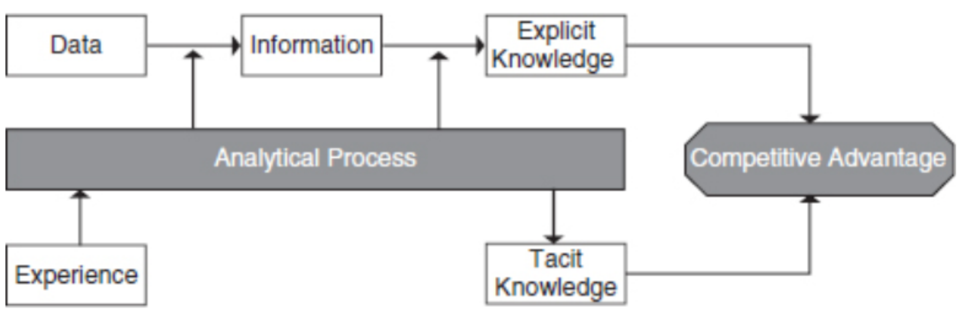

Figure 1.3 Analysis of Data as a Knowledge Creation Conduit

Theory or Data Driven?

How do we know that we know? Ascertaining the validity and reliability of what is considered "knowledge" is probably one of the oldest controversies in science. At issue here is not only the availability of objective evidence—i.e., knowledge—but also its believability. In other words, how do we separate facts from fiction? Consider the so-called "Galileo Affair"[9] to appreciate the potential difficulty of distinguishing between objectively verifiable facts and subjective beliefs. The Galileo Affair is anchored in the idea of empirically testing the key postulates of the heliocentric system proposed by Copernicus. As the inventor of the telescope, Galileo was the first to be able to gather closer celestial observations and more precise measurements empirically verifying the accuracy of Copernican thesis. Yet lacking an appreciable understanding of optics, many of Galileo's contemporaries were skeptical of his findings, suspecting the apparent celestial phenomenon to be tricks of the lenses. And so in their eyes, a great scientist was a heretic.

Obviously, our understanding of the world in which we live has increased immensely over the nearly four centuries that elapsed since Galileo's struggles. Yet as we continue to push the limits of our knowledge, we continue to struggle with the same basic task of being able to differentiate between objectively verifiable knowledge and subjective beliefs.

From the standpoint of philosophy of science, the creation of objective knowledge can be either *theory* or *data laden*. The former is carried out by means of hypothesis testing, which is a method of empirically assessing specific theory-derived claims. It means that theory laden knowledge creation springs from the foundation provided by a particular conceptual and rational paradigm—and, it is limited to testing specific claims of predictions derived from that framework. It also means that results of empirical analyses are interpreted within the confines of the theory being tested. There is a certain amount of intuitive appeal associated with that stance, but more importantly, it is supported by neuroscience studies detailing the "mechanics" of the functioning of the brain. These studies suggest that cognitive processing of information requires that our brains hold some beliefs about reality, as absent those, we would be unable to learn from the available information.[10] In other words, the human knowledge creation process appears to be somewhat confirmatory, i.e., theory laden, in nature.

In the way of contrast, the *data laden* approach assumes no existence of an underlying theory and instead approaches the analysis of the available data as a purely exploratory endeavor. Hence the resultant knowledge is built "from scratch," as it is cobbled together from individual insights found while exploring the available data. The data laden approach certainly has merit, but at the same time it is more likely (than the theory laden approach) to be blotted by data imperfections, which is a significant drawback, both in science and in business.

Business analysts need to carefully consider the strengths and weaknesses of these two competing knowledge creation frameworks. The theory laden method requires the availability of a somewhat "mature"—i.e., testable—theoretical framework. It is important to keep in mind that a handful of loosely stated suppositions or insights generated by past analyses do not necessarily constitute a "theory," at least not one that can offer an adequate explanation and/or prediction of the phenomenon of interest. On the other hand, often there can be multiple competing theoretical frameworks available, which can be confusing, to say the least. This brings us back to the earlier-made distinction between "purposeful" and "incidental" knowledge. Directing knowledge creation efforts toward clearly stated objectives, analysts can usually take advantage of conceptually robust and empirically validated conceptual frameworks emanating from a long list of business-related disciplines, such as psychology, economics and neuroscience. In fact, the marketing database analytical framework outlined in this book makes a heavy use of several such frameworks, most of all the *persuasion theory*, the *theory of reasoned action* (both "imported" from psychology) and the *equimarginal principle* ("borrowed" from economics).

The challenges faced by the data laden approach to knowledge creation are more formidable, which is a result of the combination of heightened dependence on the accuracy of data associated with this approach and the inherent imperfections of most databases. As it is intuitively obvious, greater dependence on open-ended exploration of data will create significantly more stringent data quality requirements. In purely exploratory (i.e., open-ended) data analyses, the absence of a supporting theoretical framework makes the task of differentiating between actual and spurious relationships particularly difficult, at times even impossible. Hence the question asked earlier—how do we know that we know?—becomes particularly difficult to answer.

Especially important to the assessment of the overall informational correctness of data are the notions of quality and representativeness. The *quality of data* manifests itself through data accuracy and completeness. The former represents the degree to which the coded values comprising a dataset are error-free, while the latter communicates the availability of actual (as opposed to "missing") data values for all records contained in a file. *Representativeness of data*, on the other hand, is an outgrowth of the inherent "incompleteness" of virtually all business databases. Even the largest data reservoirs contain, in effect, only a subset of all events of a particular type (such as marketing promotions), which then raises the question: How representative is a particular sample of the overall universe? However, unlike the previously discussed data quality, which tends to limit the reliability of overall findings, the representativeness of data plays a role in the generalizability of insights.

All of these considerations can be distilled to a few key takeaways. First and foremost, of the two competing knowledge creation approaches, *data laden knowledge creation* is significantly more demanding in terms of data quality and representativeness. Hence, relying on this approach to the creation of business knowledge will result in a higher likelihood of arriving at erroneous conclusions, when the available data is imperfect either in terms of quality or representativeness. This limitation is well illustrated by the drawbacks of automated data mining applications frequently used in conjunction with large transactional databases. It is one thing to flag statistically significant effects, it is yet another (and a significantly more difficult task) to reliably differentiate between persistent relationships and spurious ones.

Theory laden knowledge creation is certainly not immune to the dangers presented by poor data quality, but it is significantly less impacted by it, for a number of reasons. First and foremost, virtually no theory is ever validated or refuted on the basis of a single test, or even a handful of tests, which obviously reduces the danger associated with a single, low quality dataset. Secondly, a single study will typically test multiple hypotheses derived from the theory of interest and, under most circumstances, not all of the hypotheses being tested would be impacted in the same fashion by data imperfections.

Perhaps most importantly, theory laden approach takes fuller advantage of the cumulative nature of knowledge creation. It offers a way of leveraging previously uncovered and validated insights and building on that base with the help of additional analyses. Overall, it offers a more effective way of distilling the available data into competitively advantageous insights.

Knowledge as a Strategic Asset

An organization that "knows" how to produce something at a lower cost than other producers will enjoy cost advantage over its competitors. If its goal is to grow market share, the cost advantage will give it the ability to do so without sacrificing profitability, or it can simply enjoy higher margins and use the resultant profits elsewhere. Either way, being able to make the same product for less is clearly advantageous—does the same reasoning hold true for promotional activities? I believe so.

As noted earlier in this chapter, much of the decision-aiding knowledge comes from insights derived from the so-called "transactional" data, which encompasses a wide range of business-to-consumers and business-to-business interactions. A *transaction* is an exchange, agreement or communication carried out between separate entities and, typically, involves flow of items of value, such as goods or services. In the database sense, a transaction is an electronic record encapsulating the key descriptive characteristics of an exchange, an agreement or a communication. To comprise a distinct record, a *database transaction* has to be complete, distinct and permanent.[11]

The widespread adoption of electronic transaction processing (ETP) systems mentioned earlier has led to transactional data becoming ubiquitous over the past couple of decades. Fueled by rapid advances in computing and data storage technologies—and more recently, an explosive growth of online commerce, the ETP systems now generate enormous amounts of relatively accurate and detailed business transaction-describing data.

However, it is important to keep in mind that transactional data is outcome-oriented—stated differently, it "speaks" to what happened, but not to why it happened. This is a particularly crucial distinction from the standpoint of marketing management, best understood in the context of "chance" vs. "reason." Consider an abnormally strong performance of a particular marketing campaign—was it due to factors best described as random chance (e.g., it so happened that due to unforeseen supply chain problems, the key competing product was temporarily out-of-stock in major markets), or was it a manifestation of a particularly sound promotional strategy (e.g., the "right" offer delivered to the "right" audience)? To answer this question, one has to look beyond the realm of single-source data.

Multi-Source Analytics

It is intuitively obvious, if not altogether axiomatic that combining multiple sources of data will lead to deeper insights. From the standpoint of knowledge creation, the advantage of integrating different sources of data is the possible deepening of the understanding of the root causes of observed buyer and/or prospect behaviors. Whereas individual data silos—such as buyer behavior or demographic, psychographic and other metrics—enable one to examine the different aspects of what jointly comprises "consumer behavior," fusing together those stand-alone data sources will enable viewing consumers as integrated decision-making systems.

An *integrated decision-making system* is an analytic conception of consumer decisioning looked at from the standpoint of the available data. It captures the impact (on observed outcomes, such as product purchase or promotional response) of the relatively static characteristics, such as demographic descriptors, as well as situational factors, such as attitudinal states. In the analysis sense, it also supports the estimation of the impact of individual drivers of behavior viewed in isolation from other factors and in combination with the other factors.[12]

Integrated system analytics has been employed in a wide variety of fields, including meteorology (e.g., to analyze violent storms), geosciences (e.g., to map shorelines), medicine or aerospace design. That said, however, fusing together different sources of data with the goal of attaining deeper understanding of outcomes of interest is, under most circumstances, operationally challenging for a number of reasons, including:

- **Variability in precision.** As will be explored later (particularly in the *Data Basics* chapter), data sources typically vary—often quite significantly—in terms of their measurement accuracy. For example, due to the "what" (product purchases) and "how" (machine scanning) it captures, UPC scanner data is relatively accurate; at the same time, the often-used consumer geodemographics (basic demographic indicators, such as age, income, education, ascribed to physically proximate groups of households, the so-called "census blocks," discussed in more detail later) are relatively inaccurate.
- **Spatiotemporal differences.** Essentially all data have been collected at a particular place (the "spatio" part of "spatiotemporal," from the Latin term "spatium" for space) and time (the "temporal" part, from Latin "temporalis" for time) and the individual data sources will

commonly differ on either or both dimensions. For instance, purchases and attitudes will almost always differ in terms of time of capture; promotional response and purchase data will frequently differ in terms of both the place and time of capture.

- **Unit of measurement dissimilarity.** The basic organizational schema of data is a two-dimensional grid, where, typically, rows delimit individual "records" or "cases" and columns delimit individual "measures" or "variables." The data type specific "unit of measurement," which is what constitutes the individual rows in the data layout grid, may be an individual product (as is the case with retail scanner data), a consumer promotional response (e.g., direct mail response data) or a geographic area estimate (as is the case with geodemographics).
- **Attributional schema inversion.** One of the key operational cross-data source differences is the nature of "entity-to-action" ascription or assignment, which determines whether observed outcomes or attitudinal states are operationalized as attributes of entities (e.g., buyers or responders) or the other way around. This is a relatively technical concept discussed at length in the *Data Basics* chapter—for now, let it suffice to say that fusing together different sources of data is contingent on the individual data sources boasting comparable attributional schemas.
- **Data imputation strategy.** Once combined, multi-sourced data will often amplify individual data sources' missing value problem. Virtually all statistical analyses require an a priori (i.e., prior-to-analysis) resolution of the missing data challenge, which under most circumstances can take either of the two routes: 1. elimination of records containing missing values, or 2. imputation of missing values (both discussed in more detail in the *Analytic File Creation* chapter). Although at the level of an individual data source finding an acceptable solution to the missing value problem is rarely exceedingly vexing, finding a common-to-all (data sources) solution may be significantly more difficult.

The above outlined potential difficulties associated with the pursuit of multi-source marketing analytics should not be overlooked or underestimated—that said, the potential informational benefits are simply too substantial to ignore. As noted earlier, fusing together different sources of consumer data is a necessary prerequisite to transforming generic (i.e., available to all) information into unique (i.e., available to only a few, or even just a single organization) decision-aiding knowledge.

An old Taoist proverb teaches that "*a journey of a thousand miles begins with a single step*"; here, the journey toward the development of a robust marketing analytical capabilities starts with an explicit recognition of the tacit differences separating raw data, information and finally, decision-aiding knowledge.

From Data to Information to Knowledge

Although often used interchangeably, the notions of *data*, *information* and *knowledge* all convey fundamentally different meanings. From the standpoint of decision making utility, *data is potential information*, *information is potential knowledge*, and *knowledge is potential competitive advantage*. Implicit in these differences is a tacit value creation progression, where the initially low value *raw data* is being transformed into higher value *information*, and ultimately into the "finished product," which usually takes the form of competitively advantageous *knowledge*. Hence the notions of data, information and knowledge are all linked together by the value-adding transformation of the low impact, generic commodity (data) into high value, decision-aiding corporate asset (knowledge). The cross-firm invariance in informational proficiency is in itself a manifestation of differences in firms' analytic capabilities. An analytically proficient organization has the requisite skills and processes allowing it to reliably and consistently turn the competitively generic data into competitively unique and advantageous knowledge, while an analytically deficient organization lacks either the necessary skills or processes. Hence the attainment of informational proficiency is rooted in the development of a *knowledge creation process* as the primary conduit to the establishment of a fact-based decision making framework.

Within the confines of business—and specifically, marketing management, the knowledge creation process is defined as a *set of operationally clear data analytical activities aimed at extracting unique, competitively advantageous insights out of the otherwise generic raw data*. The operationally clear data analytical activities are further defined as *specific statistical techniques or computational algorithms*[13] *offering the most effective and efficient means of transforming specific types of data into well-defined decision inputs*. As implied in the above definitions, the data analytical process makes an extensive use of a variety of quantitative techniques, with the goal of bringing about edge-producing insights, not readily available to others in the marketplace. Culled together into a single reservoir, the collection of these insights comprises the *organizational knowledge base*, which is an increasingly important component of organizational equity. As it is intuitively obvious, the quality of the organizational knowledge base is highly dependent on the robustness of the process producing it, which as discussed later, is one of the primary drivers of cross-firm informational inequalities. Perhaps less intuitively obvious is that the ultimate measure of its efficacy is the degree to which the results attributed to that knowledge contribute to the establishment or maintenance of sustainable competitive advantage.

Some of the most revolutionary and admired companies, including General Electric, Walmart, Microsoft, Google or Amazon can attribute much of their success to informational proficiency and the impact it has had on their decision making. Although operating in vastly different industries and governed by considerably dissimilar operating models, these organizations nonetheless have one thing in common: They excel at extracting knowledge out of data and using it to their advantage. In fact, in most industries, knowledge leaders are also performance leaders, because knowing more means making better decisions in anything from resource allocation to opportunity identification. It is important, however, to consider their informational proficiency in the confines of their respective industries to account for cross-industry data inequalities. Some sectors of the economy, such as retailing, financial services or hotel and hospitality are transactional data-richer than some other ones, such as energy or materials. (This difference is primarily due to the dominant transaction type distinctiveness—e.g., retail entails business-to-consumer sales of specific items, while the materials industry is most often characterized by bulk business-to-business sales.)

In addition to data availability inequalities, the individual segments of the economy are also characterized by dissimilar levels of competitive intensity. Retail, hospitality, consumer package goods, financial services or gaming and entertainment sectors tend to be among the most competitive, in terms of the sheer number of firms offering directly substitutable products. Operating in more competitively intense environments results in firms having a greater incentive to invest early and invest more in data-supported decision-aiding infrastructure. Not surprisingly, the best known and probably the most compelling and talked about examples of unique knowledge-driven competitive advantage come from those industries. Household names, including Walmart, Capital One, Proctor & Gamble or Marriott became recognized performance leaders in their respective industries as a direct consequence of first becoming data-based knowledge leaders. These companies had the foresight to invest in the development of superior business intelligence systems and also had the discipline to make objective information the bedrock of their decision making. They have been able to consistently outperform their competitors because they are in a position to better read the marketplace and make more effective use of their resources. To them, as well as a host of other, analytically advanced data users, informational proficiency simply diminishes the amount of guesswork in such critical areas as pricing, merchandising, promotional allocation or new product design, which offers ample competitive cushion to knowledge-enabled firms.

It is remarkable that in spite of the compelling evidence pointing to significant competitive benefits associated with superior, fact-based knowledge, so many organizations continue to bet their future on intuition of individual decision makers. Obviously, there is value in the accumulated experience, but its impact can be considerably more pronounced when coupled with broader learning stemming from objective data. The goal of building a robust organizational knowledge base is not to replace the decision maker, but rather to enhance his/her efficacy by systematically reducing the level of uncertainty inherent in virtually all decisions. Knowing more enables one to spot and take quicker and fuller advantage of emerging marketplace opportunities and it is well-known that the "first mover" advantage often translates into higher profits (which is an obvious consequence of no or very few competitors). And last, but certainly not least: *Superior knowledge is also more difficult to replicate by the competition than other sources of competitive advantage*. Successful products can be copied relatively quickly, just as eye-catching promotional campaigns can be mimicked and winning strategies imitated, but because superior information is the "invisible" force behind better products or more effective campaigns, it is extremely difficult for others to replicate. Not surprisingly, firms that mastered turning data into better decisions continue to outpace their peers.

Sources of Proficiency

Informational competency is rarely, if ever, a product of an accident. Its genesis can be usually traced to careful planning of data capture/acquisition, strong execution of data analytical strategies and disciplined, system-wide embrace of fact-based decision making. It all translates into a well-defined set of "hard" and "soft" assets. In terms of the former, superior knowledge requires robust computer hardware and software, both of which are needed to support data capture or acquisition, compilation and the initial processing of the accumulated data as well as the subsequent in-depth analyses. The latter entails the availability of an appropriate data analytical skill set, without which, even the "best of breed" hardware and software will not catapult the organization to informational competency. In other words, it is entirely possible that an organization could make substantial investments in information technology and still only keep up with the competition. That is because getting ahead of the competition in terms of decision-aiding knowledge—i.e., gaining and sustaining informational, and ultimately, competitive advantage—requires looking beyond the "common-to-many" data analytic mindset in search of unique, advantage creating insights. It means pursuing a more ambitious, forward-looking informational vision.

But what does a "forward-looking informational vision" mean? In broad, conceptual terms it amounts to looking at a right problem in a right way. In more precise analytical terms, it is simply the creativity surrounding the analysis of the available data.

Knowledge leaders work at being at the forefront of informational proficiency by molding data to questions posed by critical business challenges, rather allowing computational convenience to dictate what type of information is extracted from the data on hand. They seek specific, decision-aiding insights into such key competitive-edge-producing problems as quantification of incremental sales or revenue impact of competing price and promotion strategies. They understand that in many regards, the pursuit of business knowledge is the antithesis of mass-producing generic reports that capture every imaginable nuance and detail contained in the raw data, while answering few if any of the outcome-deciding questions. That is why organizations that ultimately become performance leaders have the drive, conviction and skills to enable them to leave the comfort of the "tried and true" traditional—i.e., generic—data reporting and analyses to look for answers not yet found by their competitors. They are not satisfied with knowing as much as their competitors—instead, they search for unique, competitive-edge-producing insights. That, in a nutshell, is the essence of a "forward-looking informational vision."

Similar to other mold-breaking behaviors, analytic innovation has its own share of impediments that need to be overcome. Probably the most significant, at least from the behavior-changing standpoint is the organization's ability to sharpen its informational focus. In short, there is a fundamental difference between the *availability* and *applicability* of information. This distinction is important when thinking about data analysis (i.e., information creation) vs. data usage (i.e., information deployment). In principle, edge-producing analytics entail the inclusion of all available data, but the effective use of the resultant insights is contingent on focusing on only the subset of all available knowledge that is directly related to the decision at hand. Quite often, to know more at the decision time demands "setting aside" the bulk of the available information. Frankly, this is among the reasons that broad-base reporting tends to be an ineffective decision aid. Organizations that become knowledge leaders cultivate not only robust knowledge creation, but also rational and disciplined information usage guidelines. In a nutshell, making better decisions calls for specific, decision-related set of insights—everything else, interesting or not, is superfluous to that process.

It is a lot easier said than done, though. Many organizations' MIS/IT functions are permeated by the "volume production" mentality, which usually manifests itself in a string of detailed reports that overwhelm most, while informing only a few. It is the unspoken, though enduring belief that generating large volumes of information represents a "return" on the often hefty data infrastructure expenditures. In one form or another, the emphasis on volume is quite common, which is in a large part due to the often significant divide separating those who create information from those ultimately using it. This *analyst–user divide* is one of the reasons that the often significant investments in database and reporting technologies rarely translate into noticeable marketplace benefits. Stated differently, technologically advanced database management systems (DBMS) can be a source of *informational parity*, but not of *competitively advantageous* knowledge. This is a critical distinction and one of the key reasons behind the development of the analytical process outlined in this book.

In a strange sort of a way, the trend toward DBMS application standardization is to some degree "responsible" for the low business impact of many of these often pricey data management systems. The ever more robust capabilities of the off-the-shelf applications along with their progressively greater ease of usage, vis-à-vis the mounting technological challenges of the custom-built data processing solutions[14] have led to the virtual disappearance of the latter. Though otherwise a very positive development, the widespread adoption of generic decision support systems also carries with it some not-so-positive consequences, such as the informational conversion. Simply put, similar (at times, identical) data combined with generic data management and analysis systems often lead to very similar informational bases, quite unintentionally "shared" by organizations competing in a given industry. In the end, multiple firms competing for more-or-less the same customers and offering functionally very similar products often end up trying to outwit each other with the help of fundamentally the same information.

Still, as mentioned earlier, a relatively small but highly successful segment of companies have found a way to consistently extract competitively unique insight out of the otherwise generic data.

Their data and data management systems are usually quite similar to their competitors', but their knowledge creation is typically far ahead of the rest. The key to those organizations' success lies in how they approach the task of mining data. Unlike their less able peers, the analytically proficient firms tend to look beyond the *retrospective* (i.e., a detailing of past results), metric-centric report generation, instead focusing on *prospective* (projections supporting future decisions), business issue-centric and decision-directing insights. Reflecting fundamentally different informational paradigms, the retrospective and prospective data analytical postures differ on a number of key dimensions, with the two most important being the *degree of informational specificity*, or the *volume of the resultant information*.

It is axiomatic that the more tailored the information is to the business problem at hand and the specifics of the organization, the more it will benefit the organization's decision making. Particularly, in order for data analyses to make positive contributions to the firm's success, its outcomes have to be objectively evaluative of competing courses of action. This means that the resultant knowledge should be sufficient to point the decision maker in the direction of the greatest anticipated benefit. Although it seems like a simple enough task, making this conceptual goal an operational reality can be complex. The reasons behind the complexity are rather apparent: Many of the "traditional" data analytic techniques are ill-suited to the realities of modern transactional databases and the informational demands of the fact-based decision making. As shown throughout this book, some of the more basic techniques, such as statistical significance testing or impact (i.e., lift) quantification, are at odds with the intended use of the resultant insights—in other words, usage limitations imposed by computational processes conflict with the intended or desired business applications. Some other techniques, such as experimentation, are not per se at odds with business applications, but their usability limits are often outstretched by informational demands placed on them.

The area demanding perhaps the most fundamental re-evaluation is the broadly defined results or outcomes reporting. Virtually all organizations rely on tabulating and trending their period-by-period revenue performance, which is often "sliced and diced" in a myriad of ways. Although it is certainly important to keep abreast of historical performance metrics, this information is of little help in forward-looking decision making. Choosing among competing courses of action, such as different promotional allocation schemas, requires the decision maker to be able to quantify the schema-specific expected impact to be able to estimate their respective returns on investment, yet such insights cannot be discerned from aggregate performance results. Stated differently, making better choices demands objective estimates of *action-attributable incrementality*, which is the ultimate measure of the *worthiness* of competing alternatives. The striking dissimilarity between the basic result reporting and the objective lift assessment becomes even more pronounced upon a closer examination of both.

Action-Attributable Incrementality

The basic sales reporting is best exemplified by what has come to be known as *management dashboards*, which are reader-friendly, graphics-intensive[15] summarizations of key performance indicators. These reports are typically built around outcome tabulations, such as sales or promotional expenditures and side-by-side comparisons, where the selected metrics are broken down by type or geography. Management dashboards are clearly beneficial to decision makers insofar as they can—if well designed—present a relatively comprehensive snapshot of the business, in a highly parsimonious format. The vast majority of dashboards, however, tend to be inconclusive. They focus on easy to measure outcomes, such as sales, revenue or expenditures while failing to address the underlying *causes*. This is a considerable limitation. A brand commonly has multiple promotional vehicles at work at any given time, thus knowing which—if any—of those mechanisms had a measurable impact on the outcome of interest is quite important to future marketing mix planning.

To be more beneficial to the promotional planning process, management dashboards should be built around *action-attributable sales incrementality* assessment. In terms of the previously mentioned knowledge creation continuum, the sales incrementality assessment provides decision makers with future action guiding knowledge by linking specific outcomes with the most pronounced promotional causes. It is, however, considerably more complex methodologically. It also represents a shift in information generation philosophy: The traditional "observable outcome oriented reporting" is focused on churning out reports encompassing all that *can be known*, while the "action-attributable sales incrementality focused reporting" advocated here is focused on specific insights into *critical to know* areas.

In practice, the difference between these two methodologically and substantively distinct result-measurement approaches translates into two fundamentally dissimilar sets of activities: The former are typically built around simple metric-by-metric tabulation and summarization of aggregate results, while the latter emphasize the translating of large volumes of either inconclusive or incomplete pieces of data into probabilistic cause–effect estimates. From the analytical standpoint, the observable outcome-based reporting is computationally straightforward and nowadays the task is almost always handled by highly automated, functionally elegant data reporting software packages. The opposite is true for action-attributable incrementality assessment-based reporting, which demands highly involved, often complex and manual statistical modeling preceded by a considerable amount of data preparation. Although some parts of the requisite process are prone to standardization (hence this book), action-attributable incrementality estimation is considerably more effort and expertise intensive. That said, as exemplified by Dell, Amazon or Procter & Gamble in addition to the abovementioned Walmart, Capital One or Harrah's, the payback on the more involved knowledge creation process investment can be both significant and long-lasting.

Volume

The volume of information also matters. Interestingly, more often than not there tends to be an inverse relationship between the sheer amount of information and its business utility. This may sound counterintuitive, particularly considering our enduring belief that it is better to know more rather than less. The pitfall in this line of thinking, however, is the implicit assumption that more information translates into deeper knowledge. This is simply not true, mainly because a considerable amount of the available information often has very little to do with a particular business decision (or any decision, for that matter). For instance, the decision of which of several competing promotional alternatives should be selected when the end objective is that of profitable sales growth will not draw much help from anything other than an objective assessment of the individual promotion's anticipated lift. Other insights, such as last year's sales trends or cross-outlet sales distribution comparisons might be deemed "interesting," but ultimately they offer little-to-no help in identifying the most effective of the promotional choices. In principle, one can have lots of information but little-to-no knowledge. It follows that to yield the maximum benefit, information creation needs to be rooted in the give-and-take considerations of quality over quantity and need-directed problem solving over propagating spurious informational details. Frankly, it is counterproductive to disseminate interesting but not decision-aiding informational tidbits, simply because processing it takes time and attention away from the task at hand, while making no substantive contributions to decisions at hand.

From the viewpoint of science, there is nothing new about the notion of less being more. As one of the key tenets of scientific inquiry, this idea traces its philosophical roots to Occam's Razor,[16] a centuries-old axiom guiding the process of scientific theory building. Better known as the *principle of parsimony*, this basic “keep it simple” prescription tells us that the best explanations are those that involve the fewest number of concepts or informational details. A business decision maker will benefit much more from a relatively few, but highly action-directing insights than from a large number of mostly inconclusive and often unrelated details.

This is not to say that whatever information is not applicable to an issue at hand should be discarded—quite to the contrary, more effort should be put into cataloging and meta analysis as the means of making effective use of all informational assets. Of the two, the latter can be particularly beneficial, though it is rarely used in applied business research. Operationally, *meta analysis* is data about data, which in essence is the exploration of results of data analysis in search of underlying patterns and other communalities. It can be particularly beneficial to database analyses because it offers a robust and an objective method of summarizing the often quite voluminous basic insights. It can serve both as the means of uncovering of new insights as well as succinctly communicating of the otherwise excessively detail findings.

Clearly, informational excellence can be quite beneficial to firms’ economic well-being which brings us to an obvious question: What is required of an organization for it to develop a high level of informational proficiency? Or, stated differently: What separates informationally advanced organizations from their less analytically proficient competitors?

Analytics

Informational competency is rarely, if ever, a result of data access inequalities. As previously discussed, most organizations today have access to more-or-less that same type of raw data, largely due to the fact that the vast majority of it comes from functionally generic sources.[17] For instance, the bulk of Walmart's data comes from its point-of-sale systems which are functionally quite the same—i.e., capture the same type of basic data—as are those used by K-Mart or other major discounters (after all, these are standard applications sold and developed by outside vendors). In that sense, Walmart's informational superiority should not be attributed to (raw) data inequalities, as the previously noted standardization of electronic transaction processing systems just about guarantees that most competitors in an industry will have access to the same types of transactional data.

That is not to say that there are no instances of firms enjoying access to unique sources of data. There are a number of examples of organizations systematically supplementing their “generic” data sources with additional, often somewhat idiosyncratic (i.e., decision type-specific) data types. For instance, Harrah's enriches its basic hotel, casino and entertainment folio details with in-market experimentation results, which are then used extensively as the basis for making promotional and other decisions. The data associated with such experiments, exemplified by total dollar amounts and visit frequencies attributed to different combinations of incentives, requires specific and often complex setup and administration procedures. It means that in contrast to the “generic” transactional details passively collected by the POS and other systems, the capture of the “special purpose” data is contingent on highly specialized skill set. This means that organizations pursuing the collection of such data are typically already far more informationally proficient than their competitors. In other words, the capture of the "special purpose data" is more a result of informational competency than its precursor.

Another possible source of informational proficiency could be the broadly defined data infrastructure, which is comprised of data storage as well as data processing hardware and software. Probably even more than data access, the data processing infrastructure is extremely unlikely to be the source of informational advantage for two basic reasons: First, over the last several years the widespread standardization of warehousing and reporting applications has led to a certain degree of functional conversion, which means that differently branded applications are nonetheless quite similar in terms of their capabilities. Secondly, even the largest organizations more-or-less abandoned the earlier trend of developing from scratch their uniquely own (and thus different) decision support systems in favor of standardized, outside vendor-supplied solutions. In other words, there is little-to-no cross-user infrastructure differentiation.

In the end, neither the mere access to raw data, nor the availability of a "state-of-theart" data processing infrastructure are likely to be a source of a sustainable informational advantage. This leaves only two other plausible explanations: 1. organizational culture, and 2. the data analytical *know-how*.

Culture, defined here as the institutionalization of fact-based decision making, holds quite a bit of intuitive appeal as the source of the cross-firms knowledge disparity. After all, if an organization does not value information, in the sense of embracing data-driven decision making, it could not possibly develop a superior data-based knowledge foundation. However, one could also argue the flip side of this reasoning, namely, that it is unrealistic to expect a rational firm to value anything prior to the existence of convincing evidence. Since both sides of this argument have merit, this has the characteristics of the proverbial "chicken and egg" circular argument, with no clear way of settling "which came first." However, looking to organizational culture as the source of informational proficiency implicitly assumes that the organization has the skills required to extract uncommon insights out of the otherwise common data. And this indeed could be the crux of the problem—many firms do not.

To put it simply, the biggest single source of informational advantage is the superior *knowledge creation know-how*. Overall, the most significant factor that consistently explains why some data-rich organizations are also knowledge-rich while other, equally data-rich and technologically enabled firms are comparatively knowledge-poorer is the advanced data analytical skill set of the former. At the time when data is ubiquitous and the basic data processing increasingly informationally generic, it is the ability to go beyond the basic data crunching functionality that is the key determinant of the value of the resultant information.

Though manifestly important, the knowledge creation know-how is arguably the least developed and certainly the least formalized aspect of the new, digital world. Many will balk at this statement, as after all, quantitative data analysis itself is a well-established, long-standing field of study. And indeed it is, in the academic sense. However, as shown throughout this book, it is not in the practical business sense. Similar to a number of other fields of study, quantitative methods tend to be inwardly oriented and primarily focused on methods, rather than outcomes. Those trained in it tend to acquire substantial amounts of domain-specific knowledge, but very little understanding of the contextualizing influences of different data types or business objectives. Analysts' understanding of even the most rudimentary characteristics of modern business databases tends to lag far behind their comprehension of the specific inner-workings of the individual quantitative analysis methods, which is to some degree a reflection of many academics' limited exposure to the more complex business data sources. Frankly, that poses a problem as extracting unique and competitively advantageous insights is as dependent on the in-depth knowledge of statistical techniques as it is on the comparable knowledge of data. In short, database analytics may have the makings of a new discipline, which though heavily rooted in statistical data analysis nonetheless draws heavily from a host of other business disciplines, ranging from database technology to economics.

The Knowledge-Based Decisioning Imperative

One of the most persistent indicators of the informational maturity of organizations is their outlook on data, and particularly sales transactions. Though virtually all firms recognize the importance of this source of information, not all take a full advantage of its informational content. To the vast majority, transactional data needs to be *reduced* to be of value, which means tabulated, summarized, described and distributed via a wide range of reports, such as those showing total sales, sales per geography, time period, etc. On the other hand, a relatively smaller set of organizations take a far more *expansive and exploratory* approach to transactional data. Their more inquisitive stance stems from the desire to understand the causes of the observed outcomes, rather than merely tabulating unexplained results. Not surprisingly, their data mining capabilities are, almost always, far more developed, particularly in the sense of a wide-scale (i.e., organization-wide) use of multivariate statistical analyses and experimental design. These are the knowledge leaders discussed earlier—the organizations whose “data crunching” capabilities evolved beyond the often pointless report propagation in search of unique, competitive-edge-producing knowledge.

Interestingly, both types of organizations, the causal knowledge seekers and the result summarizers tend to speak of data as a corporate asset. In case of the former, exemplified by firms such as Capital One, Marriott, Proctor & Gamble or Walmart, data is clearly a corporate asset—after all, those firms were able to gain and maintain competitive advantage through an innovative and an effective use of data. Looking at data “asset-worthiness” through that prism, it is hard to see how the latter category of companies, the result summarizers, can make the same claim. If the firm's competitive position has not been enhanced by data, is that data really an asset to the organization? Probably not.

The reason for that is simple: Data, as a digital representation of certain outcomes (e.g., purchases) or states (e.g., demographics) is merely a raw material with a *potential* to inform the firm's decision making. Absent the know-how necessary to extract competitive-edge-producing insights out of it, raw data offers little-to-no utility to an organization, in addition to which, its informational (and any monetary) value diminishes over time.[18] For instance, 10-year-old product purchase details offer little in a way of insight into present-day product repurchase propensity or price elasticity. In other words, virtually all business data have a certain period of applicability, beyond which its informational contents become simply too dated and in effect, obsolete. At the same time, just “having” data (i.e., its capture, storage and ongoing maintenance) can be quite costly, often requiring millions of dollars of capital expenditures on computer hardware and software, not to mention dedicated stuff (database administrators, programmers, etc.). These considerations point to the question of business value: If the data residing in our IT systems does not make clear and consistent contributions to sales, or other revenue-generating activities, why should it be considered an asset? After all, basic business logic suggests that an asset should not consume more than it either currently or potentially can contribute. Let's take a closer look.

Is Data an Asset?

An *asset* is defined as something of economic value that the organization owns and controls and that is expected to provide future benefits. In an investment sense, an asset increases the value of a firm or benefits the firm's operations. Although data can be viewed as having intrinsic economic value since in many instances it could be sold in the marketplace, that argument is only applicable to a certain subset of firms (i.e., retailers often sell their transactional data to outside vendors, such as AC Nielsen or IRI, who then re-sell it, typically as packaged solutions, to manufacturers) and data types. In a broader context, it is fair to say that few if any organizations would be willing to sell their product sales data or promotional response results, as any potential monetary gains would be far outweighed by the potential competitive self-hindrance. Furthermore, there are a number of regulations governing sharing of certain types of data, such as the recently enacted Shelby Act which places severe limitations on the use of vehicle registration data. All considered, outside of the data service provider industry, few companies decide to invest in data capture and its ongoing maintenance capabilities because of the expectation of deriving an income stream from future sales of that data.

Under most circumstances, the real "asset-worthiness" of data stems from its potential to improve the firm's operations through the generation of unique knowledge, which in turn can give rise to competitively advantageous decisions. This leads to an obvious conclusion that data that do not contribute, meaningfully, to the development of competitive advantage should not be considered an asset. In fact, keeping in mind the often high cost of its capture and maintenance, poorly utilized data could even be viewed as an expense from a strictly cashflow point of view. There simply is no getting around the obvious conclusion that unless properly used, data investments can lead to an economic loss when evaluated in the confines of basic cost–benefit analysis.

All considered it is then more realistic to think of data as a *potential* asset, as such a categorization highlights the importance of the analysis of the available data. This more tenuous expression of data's asset-worthiness underscores the obvious fact that without a significant amount of effort put into analytics, even the "best" data will not contribute enough to the organization's well-being to warrant an unconditional asset designation. Also, thinking of data as a potential, rather than an actual asset draws attention to the importance of taking steps to extract economic benefits out of data that are at least equal to data's "cost of ownership."

Furthermore, thinking of data as a potential asset has a secondary benefit of redirecting the emphasis away from storage and maintenance infrastructure and toward the usage. Since the 1980s, organizations across industries have been investing heavily into data capture and maintenance-related infrastructure, while dedicating disproportionately little effort and resources to data exploration. It has been estimated that approximately 85%–90% of total data-related expenditures were directed at the hardware and software infrastructure, with only the remainder going toward extracting insights out of data. In other words, only about 10¢ out of every $1 of data-related spending went toward actually making data into a true organizational asset. As a result, the well-known expression of a firm being "data-rich, but information-poor" is often quite true.

But even the 10% or so of the total information technology expenditures that in one way or another was dedicated to data exploration has not always been utilized as much as possible. Oftentimes, a good part of that spending went toward the production of generic information (e.g., the standard, measurable outcome-focused management dashboard reports discussed earlier) that could bring the organization up to the level of competitive parity, though not sustainable competitive advantage. Some of that is due to the previously discussed convergence of technological data capture and storage platforms (e.g., UPC scanners, POS and DBMS or the recently emerging campaign management and tracking systems) combined with a generic approach to data analysis, together leading to additional informational convergence. Further fueling the informational convergence is the recent proliferation of third-party analytics, or data analysis vendors offering fundamentally the same type of information to multiple competitors in an industry. Unlike the technological standardization, however, the degree of analytical convergence varies across industries, as it tends to reflect of the availability of data to vendors. Nonetheless, there is a distinct trend of relatively few, large data providers and aggregators providing informationally non-distinct analytical products and services to a wide cross-section of the marketplace.

The slow but persistent process of technological and informational convergence underscores the importance of the earlier discussed forward-looking informational vision built around analytically innovative approaches to data analysis. Raw data has the potential of becoming an asset, but its asset-worthiness hinges on the organization's analytical skills. Data is an asset to organizations that are able to systematically extract competitive-edge-producing insights out of it. To others, specifically those whose data crunching capabilities are limited to standard, off-the-shelf tools and whose informational vision does not extend beyond basic outcome reporting, data is yet another component of the cost of doing business.

Data as a Source of Competitive Advantage

The last couple of decades have been particularly eventful from the standpoint of business information. Some of the more noteworthy trends, from the standpoint of knowledge creation, include the following:

- A combination of rapid gains in data processing capabilities, decreases in storage and processing costs and the proliferation of powerful software applications are leading to database technology becoming affordable to an ever growing number of organizations.
  - Result: *Leveraging customer and competitive data became a key ingredient of firms' product and market strategies*.
- The growing digitalization of business processes, including sales and customer interactions, is spawning an overabundance of transactional data often leading to potential users "drowning in data but lacking information."
  - Result: *Organizations spent large sums on customer data warehouses, yet to-date only a handful truly leverage their data assets*.
- The forces of deregulations coupled with growing business globalization are leading to considerable increases in the level of competition, ultimately amplifying the importance of timely and accurate customer and market information.
  - Result: *Increasing competition accentuated the need for speedy extraction of actionable business insights from databases*.
- Information availability and immersion are becoming a part of everyday business culture and data analysis techniques slowly making their way into common business lexicon.
  - Result: *As database analytics is no longer a domain of a few, large organizations, the demand for skilled analysts exploded*.

Taken as a group, these developments are to a large degree responsible for the growing importance of unique (to a given organization), fact-based knowledge in building and sustaining competitive advantage. In a sense, *all organizations are now in the information business*, to the degree to which their competitive well-being has increasingly grown dependent on the access to timely and accurate decision-guiding insights. Stated differently, knowledge surrounding the key decisions, such as product design (i.e., what is the most desirable bundling of product attributes?), promotional mix allocation (i.e., how should the finite promotional dollars be allocated across the available promotional alternatives to deliver the highest incremental benefit?) or the customer acquisition and retention strategy (i.e., what are the most cost-effective tools/tactics?) is now among the most pronounced determinant of firms' success.

In a recent, insightful look at the impact that the persistent and well-thought-out data analysis—defined as reliable conversion of raw data into competitively advantageous knowledge—can have on organizations' long-term success, Thomas Davenport delineated a number of key factors characterizing information-driven organizations.[19] These include the widespread use of modeling and optimization, enterprise-wide deployment of data-derived insights and solid support from the top echelons of management. Of those, the *widespread use of modeling and optimization* comprises the general set of skills needed to translate the mounds of often dissimilar raw data into useful information. As detailed by Davenport, the quality of the resultant analyses requires the coming together of three key components: the right focus, the right people and the right technology. Implicit in the interplay of those three information-quality shaping forces is the analytic know-how, which is that somewhat intangible quality that on one hand calls for strong quantitative methodological skills, while at the same time contributing a healthy amount of problem solving creativity. A disciplined and rational left brain meets the spontaneous and untamed right brain…not impossible, but at the same time, not an everyday occurrence either.

As previously outlined in the *Sources of Proficiency* section, organizations that excel at extracting competitively advantageous knowledge out of the otherwise generic data are able to do so because of their data analytical prowess and the organization-wide fact-based decision making discipline. Processes ranging from high-level strategic planning to tactical decisions surrounding product mix, logistics, inventory and distribution or promotional mix spending allocation are all making increasingly better use of the available data. As mentioned earlier, one of the key drivers that fueled Walmart's growth and its eventual ascendance to the world's largest retailer and the Fortune #1 organization (based on gross revenue[20]) was its early embrace of the information-driven decision model. While its competitors continued their march forward (or so they believed), guided mostly by their intuition, anecdotal evidence and rarely empirically validated generalizations, Walmart looked to objective and representative data insights for decision cues. As a result, its merchandising mix consistently outperformed its peers, while its simulation- and optimization-based supply chain management mercilessly squeezed unnecessary inventory and stock-out costs, enabling it to offer competitive prices while still generating attractive returns for its shareholders. It is no surprise that the now industry-leading organizations such as Dell and Amazon emulated Walmart's supply chain philosophy as one of the engines catapulting them to the position of prominence.

But Walmart's way is not the only way—frankly, blindly copying methods of successful companies' practices can be a slippery slope, as it cannot be assumed that just because a particular practice or a method works well in one instance, it will work equally well in other instances. Inherent in the development of effective data mining capabilities is a certain element of organizational self-discovery, which entails the identification of the most adaptable (to the specific of the organization) ways the organization can use data to gain and sustain competitive advantage. After all, the retail industry's dynamics are quite different from the pharmaceutical, financial or other sectors of the economy, as is the available data. Neither are any two firms in a given industry exactly the same, particularly in the cultural sense (for instance, some firms are highly centralized while others are de-centralized in terms of their decision making models; some are overt risk takers in terms of heavy emphasis on new, trend setting products, while others are risk avoiders, preferring instead to focus on the "tried and true" ideas or technologies). Therefore, the specifics of "what" data and "how" it can be used effectively are shaped by both industry-wide forces (e.g., what type of data and data insights can offer the greatest potential competitive levers?), as well as by company-specific competencies and goals (e.g., company's intrinsic capabilities and its organizational culture).

Capital One, as one of the leading credit card issuers in the U.S. and one of the leading credit card industry innovators has consistently delivered above average results by an almost religious dedication to objective data analysis, especially customer mix optimization. Harrah's, a major casino entertainment organization and a relative risk-taker in its industry, systematically improved its profitability by using data analysis to attract and retain customers with the greatest profit potential. A resource-constrained baseball club, the Oakland A's consistently posted one of the league's best regular season records by identifying the otherwise undervalued players that could perform the desired assortment of tasks, something Oakland was able to do in spite of working with a far-below-average budget. Honda developed a second-to-none brand loyalty by using data for early detection of potential problems, thus greatly increasing the reliability of their automobiles. A leading hotel chain, Marriott, uses advanced analytics to optimize the price–profitability relationship, while Novartis, the giant pharmaceutical firm leverages data analysis to improve the quality and the efficacy of its R&D efforts. Last but not least, Procter and Gamble, a leading consumer packaged goods manufacturer continues to prosper in a very competitive, mature industry segment in a large part due to organization-wide embrace of data-driven new product development, as well as promotional mix allocations and evaluation practices.

The Emerging Field of Database Analytics

The rate at which the data is accumulated vastly exceeds the rate at which it is analyzed. The ever expanding digitalization of business transactions, manifesting itself in the growing number of data-creating technologies, such as POS systems or campaign tracking software resulted in an exponential growth in the sheer amount of raw data. In addition to the "involuntary" data capture, systematic data "creation" has exploded, both in terms of types as well as volume. Specialized databases, such as consumer geodemographics, psychographics, spending or wealth estimates are widely touted as potential sources of the "whys" behind the "whats" offered by transactional data.

Yet as mentioned earlier, on average, out of every \$1 spent on databases, no less than \$0.85 goes to storage infrastructure and only roughly up to \$0.15 is allocated to extracting insights out of the data. Some of this spending allocation disparity can be attributed to the more capital-intensive nature of data capture and storage infrastructure; however, that reasoning does not extend beyond the initial acquisition and setup costs. In other words, once the data capture and storage systems are in place (and paid for), the infrastructure vs. analysis resource allocation inequality should begin to disappear. But that rarely happens in practice, which is a manifestation of an enduring—and incorrect—organizational mindset equating "information" with "data." It is common for organizations to believe that continuing to pour money into data capture and maintenance infrastructure will lead to better decisions and, ultimately, higher revenue or profitability. The empirical evidence, however, does not support that assertion, as there is no clear cause–effect linkage between the organization's IT expenditures (defined as hardware- and software-related spending) and its marketplace success. Operating system, processor or storage hardware upgrades may enhance the organization's data capture, storage, retrieval and processing capacities, just as business intelligence data reporting software will likely increase the ease, the speed and the usability of basic database reporting. In the vast majority of cases, these are the real informational benefits of the aforementioned $0.85: more data—more accessible—more quickly.

The ever larger and more numerous organizational databases are not only expensive to build and to maintain, but can also be analytically taxing. In fact, if the data they contain is not analyzed correctly, unreliable or even invalid inferences can be drawn. For instance, it is now commonplace to rely on the notion of statistical significance to attest to effect coefficients' (such as promotional response lift) "validity," which as detailed later often outstretches the test's application limits (which is usually due to either the sheer size[21] of the analysis universe or the nature of the data,[22] both of which require controlling steps that are rarely taken in practice). Ultimately, when subjected to a closer scrutiny, a good many of the seemingly "robust" findings end up being nothing more than spurious relationships misinterpreted as facts.

The analysis of large corporate databases can also present significant interpretational challenges. The explosion in the availability of data associated with the advent of electronic transaction processing made possible an objective evaluation of a wide range of promotion-related and other types of behaviors, but doing so oftentimes necessitates the use of complex, probabilistic multivariate statistical models. These approaches tend to make heavy use of abstractly defined "statistical effects" that require a considerable amount of interpretation before becoming meaningful to non-technical business audiences. This is particularly the case with statistical models employing interaction (i.e., combinations of variables expressed as separate metrics) or non-linear (i.e., effects where the change in one metric is accompanied by a disproportionate change in another) terms. Though no one knows with certainty, it is believed that more than 50% of database modeling findings never mature beyond the initial, technically framed presentation. Hence it follows that the attainment of an acceptable level of the decision making utility is tied to the adequacy of "model translation" efforts, or the re-casting of obtuse statistical effects into more interpretationally clear business terms. And that in turn is dependent on the adequacy of the analyst training.

On-the-job training is still the most popular means of training junior analysts in working with large volumes of heterogeneous data filling up the plethora of databases found in many of today's business enterprises. Obviously, it can be very beneficial for the more experienced employees to pass their knowledge onto their less experienced colleagues; however, it can also contribute to the institutionalization of flawed practices. Moreover, it promotes "learning without understanding," a highly undesirable practice for any organization aspiring to glean competitively advantageous insights out of its data.

An obvious alternative to on-the-job training is more specialized academic preparation. Unfortunately, business schools have been relatively slow to both recognize and embrace (in the sense of teaching) database analytics. More often than not, university-level statistics courses teach the theory in a relative isolation from the idiosyncrasies of different data types and, when using data, rely on overly simplistic datasets. In the end, the bulk of the graduates end up being ill-equipped to tackle the complexities of modern business databases.

What steps should data-rich organizations take to also become information- and (more importantly) knowledge-rich? Make better use of the available data resources through the development and institutionalization of robust data analytical capabilities. Doing so, however, requires a couple of key steps: First, organizational knowledge creation needs to be approached as an ongoing process, rather than a sporadic event. In other words, an organization aspiring to become a knowledge leader in its industry needs to put in place a process linking organizational goals, informational objectives, data analyses and knowledge proliferation. Secondly, insight validation and predictive reliability processes need to be established, and need to be capable of differentiating between spurious associations and true competitively advantageous insights. This will call for the proverbial "thinking out of the box" to adapt statistical methods and processes to the characteristics of large corporate databases and the demands of business decision makers. This book offers an outline of a database analytical process, designed expressly to help business organizations increase the impact of their marketing promotions.

The Uniqueness of Marketing Database Analytics

As an emerging, practice-oriented field, database analytics draws on multiple disciplines ranging from statistics to database technology, promotional campaign management and strategic planning. As a subset of the larger database analytical process, promotion analytics relies on largely the same universe of skills and knowledge, with a particular emphasis on the following:

Technical Skills

- **Statistics.** Database analysts are expected to be proficient in a wide range of the univariate and multivariate techniques, the sampling theory and the experimental test design.
- **Campaign measurement.** Probably one of the least developed areas within the field of database analytics, it focuses on measuring sales incrementality attributable to individual promotional treatments, programs and campaigns.
- **Online analytics.** Typically focused on Web traffic analysis, tends to be somewhat idiosyncratic in terms of tools and to a lesser degree, methodologies.
- **Database technology.** Although database analytics practitioners do not need the level of technical proficiency of developers or programmers, knowledge of the basic level of relational or event database structure is important to make the most of the data they contain—i.e., to translate the often voluminous data into sustainable competitive advantage, which is the theme of this book.

General Promotion Knowledge

- **Promotional strategy, targeting and segmentation.** The knowledge of the basic conceptual frameworks along with practical applications.
- **Campaign design and management.** The basic flows and processes; differentiation between offer, treatment and campaign.
- **Direct marketing.** The basic design, fielding, tracking and measurement techniques.

Experience-Based Knowledge

- **Promotion analytics "rules of thumb."** Is 20% a reasonable campaign response rate to a repurchase campaign targeting lapsed brand buyers? Sometimes, to be able to pinpoint a truly successful result or a finding that falls outside what is normally considered to be a "reasonable" range, an investigator needs to have some degree of familiarity with "typical" or "normal" values for given outcomes.

Why is it important for a data analyst to develop such a wide-ranging set of skills and proficiencies? The short answer to this question is because all of those factors play a role in transforming generic information into unique, competitively advantageous knowledge. More on that in the next section.

The Marketing Database Analytics Process

As convincingly argued by Peter Drucker, one of the most preeminent management thinkers, a business enterprise ultimately just has one goal—to create a customer. Doing so requires management to perform two critical functions in every organization: 1. to foster innovation, and 2. to effectively market the organization's products and/or services. Management teams that perform these two tasks better than their competitors will win a larger share of the overall customer base.

It follows from the above rationale that effective promoting (of products or services) entails using the least amount of resources to win the largest number of customers. This intuitively obvious notion is at the root of what is generally referred to as "promotional productivity," often objectified through the promotional ROI (return on investment) metric.

Reaching—and more importantly—sustaining above average levels of promotional productivity requires a considerable amount of specific knowledge, such as buyer purchase or repurchase propensities, price elasticity, anticipated promotional lift, etc. That said, the seemingly long list of the potentially useful decision-aiding insights can be reduced to a relatively small set of well-defined critical promotional success factors. However, what constitutes critical success factors for marketing is in itself a function of the role of marketing vis-à-vis the organization at large.

When looked at from the standpoint of business utility, marketing promotions (as a functional area within an organization) are expected to offer a set of tools to aid the organization in reaching the stated performance objectives. In other words, promotions are means to an end, where the end represents reaching or exceeding the said business goals. And although business enterprises vary significantly in terms of what and/or how they produce, their goals can nonetheless be reduced to one or more of the following three general categories:

1. Sales/revenue growth.
2. Profit growth.
3. Market penetration.

Naturally, most organizations will be concerned with all of the above, but at any given time, any one of those goals will be more pronounced than the remaining two. For instance, a relatively new and small organization will typically emphasize sales growth and market penetration over near-term profitability, while a more mature and established firm will typically be more concerned with increasing the profitability of its operations. And ultimately, since all organizations have to make tradeoff-entailing resource allocations, the resource constraints will usually "force" them to prioritize their objectives.

Whatever the stated business objectives, the role that marketing promotions usually play in reaching those goals falls into one of the three distinct sets of activities:

1. New customer acquisition.
2. Current customer retention.
3. Customer value maximization.

Not surprisingly, larger organizations tend to have somewhat separate functions, or departments, each tasked with new customer acquisition, current customer retention as well as cross- and up-sell (customer value maximization). Whether that is the case or these somewhat different tasks are all concentrated in the hands of the same group, the role of promotional analytics is nonetheless to supply marketing personnel with actionable knowledge that can be used to enhance the efficacy of the aforementioned key promotional endeavors.

Hence it follows that a generalizable marketing database analytical process should contribute to the efficiency and the effectiveness of the above-delineated promotional objectives, and by extension, the broader organizational goals. This is particularly important from the standpoint of drawing an objective line of demarcation separating the “truly impactful” from the merely “interesting” insight-generating analytic endeavors. In other words, the exploration of the available data that is not directed toward specific informational needs is likely to produce a plethora of intellectually curious, but practically insignificant findings, all of which will contribute very little toward increasing the efficacy of promotional efforts.

The process-based approach to effective analyses of large marketing databases summarized below was developed within the constraints imposed by the above considerations. It outlines the means of translating raw data into, initially, *informational insights* and ultimately, into *competitively advantageous knowledge*. The overall process flow and its key components are depicted in Figure 1.4.

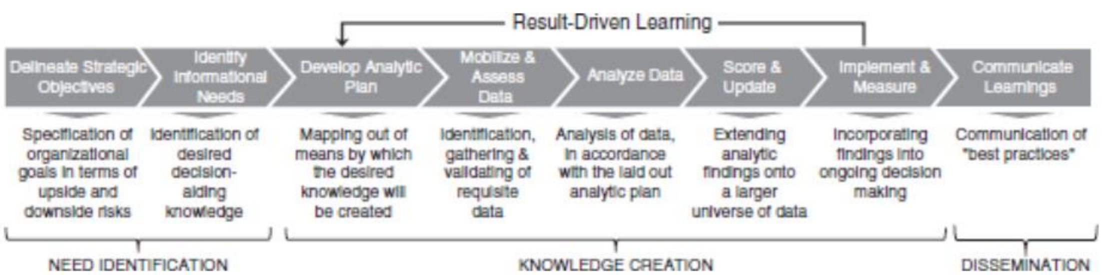

Figure 1.4 Marketing Database Analytics Process

A more in-depth discussion of the database marketing analytical process shown above is presented in the next chapter.

About This Book

This is a “how-to” book: How to use commonly available data to get unique insights into commonly asked marketing promotion related questions. Given its analytical focus, the bulk of the book's contents are related to quantitative data modeling. That said, non-statistical modeling issues are explored with a comparable level of depth and care, as the creation of truly unique and valuable business insights necessitates a solid understanding of elements of other supporting domains, such as strategic planning, data warehousing and economics.

Many of the solutions presented here stem from hands-on practical experience of working with numerous, large organizations from a cross-section of industries, including financial, energy, automotive, packaged goods, insurance, banking, telecommunications, high tech, retail, consumer durables and hospitality. The overall analytical process and its individual components are expressly “calibrated” to take into account the most commonly used data sources, with the primary emphasis on large transactional data warehouses, such as the UPC scanner product movement data typically associated with retail point-of-sale systems. Also explicitly considered are other common business sources of data including sample-based surveys, such as consumer satisfaction or product quality studies; data aggregator-compiled “causal overlays” exemplified by consumer geodemographics; public financial filings and disclosures, such as the Standard & Poor's Compustat database tracking quarterly and annual financial filings that are required of publicly traded companies. In the data sense, the basic premise of the database analytical process presented in this book is the amalgamation of dissimilar sources to form the basis for creating the otherwise-beyond-reach knowledge.

Content-wise, I draw upon the work of multiple disciplines, including statistics, business strategy, database technology and management and to a lesser degree from other areas, such as industrial psychology. In my discussion and recommendations I rely heavily on practical database analytical consulting experience, based on my work with large corporate databases of Fortune 500 organizations from a wide cross-section of industries mentioned earlier. I make extensive use of database analytical examples drawn from real-life analysis and modeling projects to illustrate the shortcomings of some methods and processes and to underscore the advantages of other approaches.

As stated above, this book is about the "how-to" of database analytics. Although it offers theoretical descriptions and rationales behind the recommended approaches, the focus throughout this text is on the hands-on, systematic process of answering specific business questions. It is important to note that, as detailed in Chapter 2, the end goal of the marketing database analytical process, depicted above in Figure 1.4, is not a mere translation of data into information, but rather it is to create a source of unique and thus competitively advantageous organizational knowledge, added to and updated on an ongoing basis.

My secondary objective is to contribute to a re-focusing of analysts' attention away from the "modeling" part of the process and toward the "usage" of the resultant knowledge. This is not to say that I am advocating modeling carelessness, far from it, as the subsequent chapters show. I am merely trying to shine the spotlight on the utility of the end product of analysis, recognizing that an abstractly defined statistical model—in and of itself—rarely offers meaningful amount of utility to its end users, unless the results are properly framed and communicated.

Organization

The basic organizational framework of this book follows the general outline of the framework depicted in Figure 1.4 above. Although the content is of relatively technical nature, the presentation is (hopefully) easy-to-follow.

The next chapter (Chapter 2) offers a more in-depth discussion of the marketing database analytical process shown in Figure 1.4. The remaining chapters are grouped into three broad sections that make up the marketing database analytical process: 1. Need Identification; 2. Knowledge Creation; and 3. Dissemination.

The first two chapters of the *Need Identification* section, *Organizational Objectives and Informational Needs* (Chapter 3) and *Skills and Tools* (Chapter 4), kick-start the marketing database analytics process by drawing attention to the importance of rooting the analysis of data in specific organizational needs, all with the goal of increasing the chances of the resultant insights making a substantive and material contribution to the marketing action-related decision making.

The second section—*Knowledge Creation*—comprises the bulk of the book's content, nine chapters in total (Chapters 5–13), as it details the individual components of the systematic transformation of raw marketing data into decision-aiding insights. Chapter 5, *Analytic Planning*, describes some key considerations entailed in laying out not only what analyses are to be done, but also any prerequisites that need to be satisfied. Chapter 6, *Data Basics*, weaves together a number of concepts from database design, information theory and database management all aimed at forming an objective foundation of the basic concepts surrounding electronic data storage and manipulation. Chapter 7, *Analytic File Creation*, offers an in-depth treatment of specific data cleansing and preparatory steps, geared toward assuring the accuracy of input data as well as the attainment of distributional and other data properties that are essential to assuring robustness of statistical analyses of data.

Chapters 8 through 11 detail the heart of the marketing database analytical process—the actual analysis of data. Chapter 8 presents a discussion of what should be the initial data analytical step—*Exploratory Data Analyses*, the discussion of which is geared toward establishing the foundation of the informational content of the earlier created data file. Building on those initial, exploratory insights, is a relatively broad *Segmentation* discussion (Chapter 9), followed by a more focused review of *Behavioral Predictions* (Chapter 10), which is built around key (to marketing analytics) multivariate statistical techniques. Next (Chapter 11) comes a discussion of *Action-Attributable Incrementality* approaches, focused