Before the Fork in the Road: Investments in Research on Competing Alternatives

Laura J. Kornish[1]

May 24, 2006

[1] Assistant Professor, Marketing Division, Leeds School of Business, University of Colorado, UCB 419, Boulder, CO 80309-0419, kornish@colorado.edu, 303-735-6385.
The author thanks John Lynch for helpful discussion and suggestions.

Before the Fork in the Road: Investments in Research on Competing Alternatives

Abstract

In many decisions, firms face binary choices: one alternative or another can be pursued, but not both. This paper examines the logic behind initially exploring multiple approaches when constraints or the structure of returns will ultimately force such a choice. Many reasons have been offered as rationales for diversification of investment, such as risk aversion and economies of scope. In this paper, I describe a new rationale for diversification, based on an uncertainty-resolution or "value of information" explanation, and derive the conditions under which it holds. If a single alternative will eventually be picked, investment in research in multiple alternatives has the greatest incremental value over investment in research on a single alternative in the following scenarios: 1) when, with joint information, a good outcome from the investigation of an alternative might nonetheless result in another alternative's being picked because the other alternative has an even better outcome, and, likewise, 2) when a bad outcome might result in a alternative's being picked because the other alternatives are even worse. Even when the outcomes for different alternatives are probabilistically independent, decisions about research cannot be made separately for each alternative; the ultimate comparison of the alternatives links the research decisions. With probabilistic dependence, if each alternative has a non-zero value of information individually, the incremental value of joint information over individual information is highest when the alternatives are mildly positively dependent.

Key words: market research, value of information, diversification, sequential investment

1. Introduction

In many decisions that firms face, constraints force binary choices: one alternative or another can be pursued, but not both. For example, a former student of mine started a network security company a few years out of business school. One of his partners had written “alpha-level” code for software for each of two different markets: wireless networks and wired networks. Both concepts were then considered technically proven, but as a start-up, they were resource-constrained, and they framed their choice as “wireless or wired?”

Their constraints on cash and manpower were typical for a start-up, and those constraints forced them to make a decision one way or the other. These are not the only types of constraints that force either-or choices. The constraint may be physical: on a particular piece of land, a developer builds either a contemporary or traditional home. The constraint may be temporal: a television network decides which show to run in a popular time slot. The constraint may be on attention: ultimately, a firm can have only a single overarching strategic direction or high-level advertising message.

Another issue that forces choice is the structure of returns. If two promising opportunities each face increasing returns to scale, then the optimal investment pattern is to pick a single opportunity. Large fixed costs (even without cash constraints) create increasing returns over a wide range of investment, and therefore dictate that a single approach be pursued.

Pringle (reporting in *The Wall Street Journal*, 2004) describes Nokia’s great commitment in the early 2000s to smart phones, “cellphones that allow consumers to get on the Internet, watch movies and play video games,” at the expense of midrange phones. Pringle reports that Nokia “has spent hundreds of millions of dollars launching a string of ‘smart phones.’ It now pours almost 80% of its research-and-development budget—about $3.6 billion a year—into software, much of it designed to give phones computer-like capabilities.” The article describes how Nokia eschewed the midrange phones and bet on the higher-end smart phones. With the large fixed costs of R&D, it made sense for Nokia to “pick a winner” rather than hedge bets by developing both midrange and high end phones. (At the time the article was written, Nokia had not, in fact, picked the winner.)

These either-or choices, which often involve large stakes, raise the motivating question of this paper: under what conditions does it pay to research competing alternatives, when a single alternative will ultimately be chosen? The contribution of this paper is to articulate a new reason for simultaneously pursuing multiple approaches. The reason developed in this paper is logically distinct from the other arguments for diversification of investment. Other, well-established reasons for diversification include risk aversion (Markowitz 1959), decreasing returns to scale (Prastacos 1983, Loch and Kavadias 2002), economies of scope (Teece 1980), transaction costs (Teece 1982), strategic barriers to entry (Bain 1956, Dixit 1979), and product lines (Moorthy 1984). My reason derives from an information-gathering story. Simultaneously pursuing multiple approaches, while costly, allows for more-informed comparisons among the alternatives, leading to a better choice down the line.

My analysis applies to a broad range of situations, including contexts as different as advertising campaign selection and product development. At the very earliest stages of a decision process, it clearly makes sense to entertain many new ideas. As projects progress, however, it becomes increasingly expensive to advance ideas, so unpromising ones must be weeded out. There is a trade-off: carrying more ideas forward is costly, but allows for the possibility that one will develop into something terrific; carrying fewer ideas is less expensive, but may kill off a potential blockbuster. The farther into the decision process you are, the tighter this tension becomes.

As mentioned above, there are many factors that contribute to resolving this tension, such as risk aversion (in which a risk-averse decision maker prefers a portfolio of negatively correlated assets to a bet on the most promising asset) and development of product lines (in which the firm launches several variations of a product to appeal to heterogeneous customers). I do not provide an exhaustive list of those other factors, but rather propose what I believe is a novel argument for diversification. I focus on the role of investments in resolving uncertainty. Investing resources in the early stage of a project serves two roles: investment is used to improve the idea (what Cohen, Eliashberg, and Ho 1997 call the “product performance improvement process” p. 118), and it is used to resolve uncertainty (e.g., learning about market response or technical feasibility).

Through the latter role, logic may dictate that multiple project ideas be advanced. The aim of this paper is to explain the conditions under which that logic holds.

I construe the information-gathering activities as belonging to the first of two stages of investment. In the first stage, investments resolve uncertainties related to the value of the alternatives (e.g., through market research). In the second stage, a commitment to at most one of the alternatives is made. To analyze the value of the first-stage information-gathering activities, I use the "value of information" concept from the economics of information and decision analysis and which is found in many marketing research texts (e.g., Aaker, Kumar, and Day 2004 and Malhotra and Peterson 2006). The value of information is the additional value of knowing the outcome of an uncertainty before having to make a decision, compared to the default case in which the uncertainty is not resolved until after the decision is made. Despite the popularity of the value of information concept in marketing pedagogy, it has proven difficult to characterize: Hilton's 1981 paper laments the mostly negative results concerning value of information and Delqiue's (2005) recent work confirms the continued difficulties.

Consistent with the negative results, it is known that the value of information on two uncertainties is not additive: that is, the sum of the value of information on two uncertainties considered separately does not necessarily equal the value of information on the same two uncertainties considered jointly. In general, the value of information on two uncertainties can be either superadditive or subadditive. (Hilton did not include this particular negative result, and there is no well-known citation for it, perhaps because the result is negative.) What we do know, from Blackwell (1953), is that there will always be more value in resolving two uncertainties jointly than in resolving either one alone, when uncertainty resolution is costless. When uncertainty resolution is costly, however, even the "two are better than one" rule does not always hold. It holds only when there is a great enough synergy (mathematically, superadditivity) in the joint information to justify the cost of resolving both uncertainties. The superadditivity arises when there is great value in making a more-informed comparison between options.

The essence of my argument, based on two alternatives, is as follows. Investigating a single alternative allows a gross comparison: how does the one under investigation compare to the mean value of the other one? If the investigation reveals favorable information, pursue the alternative being investigated, otherwise, pursue the other. Investigating both alternatives allows for a more intricate comparison. The joint information will be particularly attractive when one alternative has the best upside outcome and the other alternative has an unattractive downside. In this "favored alternative" scenario, collecting information on both alternatives allows for an outcome that the gross comparison from the information about a single alternative doesn't allow. With joint information, a good outcome from the investigation of an alternative might nonetheless result in the other alternative's being picked, if the other one has an even better outcome. Likewise, a bad outcome might result in that bad alternative's being picked, if the other one is even worse.

I show that this logic holds under both probabilistic independence and dependence of the outcomes. With dependence, I find that if both alternatives individually have non-zero value of information, then joint information enjoys the largest incremental value over individual information when the outcomes of the alternatives have a slightly positive dependence or correlation. Too strong dependence (either positive or negative) obviates the need for investing in learning the outcomes for both. Even though positive dependence increases the chance of the worst-case outcome, it provides greater opportunity for the intricate comparison that drives the value of joint information.

After reviewing the related literature in the next section, I elaborate on the logic of the intricate comparisons afforded by joint uncertainty resolution. In section 3, I build and analyze a requisite model (i.e., the simplest model that makes the point) that assumes independence of the uncertain outcomes across alternatives. In section 4, I relax the independence assumption. Section 5 covers the robustness of the conclusions to other structural assumptions in the model, and section 6 concludes with discussion and examples.

2. Related literature

In many of the contexts, my model's first-period investigations can be thought of as market research investments. Much attention has been paid to the specifics of market research techniques in the marketing literature (e.g., Anderton et al. 1980, Chatterjee et al.1988, Urban and Katz 1983, Zaltman 1997); this paper is complementary to that work. Abstracting from the details of specific investigative techniques, I explain the conditions under which considering a portfolio of research projects adds value beyond individual investigation. Similar in spirit is the work of Farley, Lehmann, and Mann (1998), who also write about the optimality of research approaches.

This paper has links to the vast literature on project selection that exists beyond the marketing literature (also referred to as R&D portfolio selection), which is spread over the operations research, management science, and industrial engineering literatures. Loch and Kavadias (2002) give a nice overview of the project selection literature. Much of this type of work is formulated as single period investment allocation decisions. Some papers, such as Dickinson, Thornton, and Graves (2001), capture the continual investment over time. The current work focuses on two types of investments that are distinct, both in time and in purpose. The early stage investments are for uncertainty resolution or learning; the later-stage investments are for the actual execution. Similar to this two-stage decision, Mehrez and Sethi (1989) study a context with multiple projects and the opportunity to purchase information. They examine the effects of budget constraints on the information-gathering phase for independent projects. This approach contrasts with mine, which focuses on the trade-off between costs and benefits of information gathering, rather than on a constraint on the activity. Like my work, a short comment by Mehrez (1989) looks at investments in competing approaches without budget constraints on the information-gathering activity.

This sequential investment orientation draws inspiration from the work on absorptive capacity: that is, the capability of firms to evaluate and capitalize on knowledge in a particular domain (Cohen and Levinthal 1994). Cohen and Levinthal develop the absorptive capacity concept to explain why firms should and do invest in currently less-promising technologies or projects. They use a framework of sequential decision making under uncertainty, as I do here. In their model, the firm can invest in absorptive capacity in the early period, and that investment will affect both its perception of subsequent noisy signals as well as its ability to profit from developments in the domain. Their model and mine both consider that investments made in the earlier period create a more favorable decision-making environment in the later period. The main difference is that in the work of Cohen and Levinthal (1994), the initial investment affects the precision with which extramural information is processed, while in my work the initial investment generates the information itself. The contribution of this work beyond the issues explored in that paper relates to the issue of the interaction of investments: While they look at investment in absorptive capacity in any one area in isolation, I examine how the presence of multiple projects influences the prescribed investment behavior.

Finally, the information-gathering perspective of the current work links it to the literature on design of experiments, especially work on optimal sample sizes in Bayesian decision theory. Especially relevant is the work on sampling to determine the best of several processes (Pratt, Raiffa, and Schlaifer 1995, ch. 23A; Raiffa and Schlaifer 1961, ch. 5B). The context of that work is similar to the current work in that the optimal sample sizes are for a single-shot experiment, not sequential, contingent experiments. An important difference is that in the current work, the alternatives include the option of investing in neither approach, whereas in the optimal sample size work, the results all depend on the assumption that the best process is always better than doing nothing.

3. Model and analysis

In this section, I present the simplest model that illustrates the main thesis of the paper, that under certain conditions, synergies in information collection make diversified investment better than focused investment. In the next two sections, respectively, I look at the issue of probabilistic dependence and examine the robustness of the conclusions.

In this model, there are two alternatives under consideration ("alternative A" and "alternative B") under consideration. I use the generic word "alternatives" to allow for the interpretation of the results in different contexts. If the context is product development, the alternatives can be two product concepts or at a higher level, two markets on which to focus. If the context is an advertising campaign or a brand message, the alternatives are the candidate campaigns or messages. If the context is real estate development on a particular plot of land, the alternatives are the home designs, or on a larger scale, the type of development, such as family-oriented resort or a high-end spa.

The model has two stages of investment. In the first stage, one can invest in these alternatives to learn about the uncertainty in their value. In the second stage, a final commitment is made; either both alternatives are abandoned or one of them is selected. The introduction mentioned several reasons for the all-or-nothing character of this second stage (increasing returns to scale, cash constraints, etc.).[2] I discuss the situation in which both alternatives can be done in section 5.

Alternatives A and B have mean values denoted a and b, respectively, such that $a > b$. Each alternative has two outcomes, a high outcome (a_1 and b_1, respectively) and a low outcome (a_2 and b_2, respectively). With two outcomes, it follows that

$$a_1 > a > a_2 \text{ and } b_1 > b > b_2. \quad (1)$$

Outcome a_1 happens with probability p (and a_2 with probability 1-p), and b_1 happens with probability q (and b_2 with probability 1-q). The means are therefore $a = pa_1 + (1-p)a_2$ and $b = qb_1 + (1-q)b_2$. In this section I assume probabilistic independence.

To demonstrate the logic of the value of diversification in information collection, I assume that both alternatives are positively valued in expectation: $a > b > 0$, and there is positive value of information for each alternative separately. The latter assumption implies that if you knew an alternative would have the high outcome, you would prefer that high outcome to the mean value of the other, and if you knew an alternative would have the low outcome, you would prefer the mean value of the other alternative to that low outcome. In my notation,

$$a_1 > b > a_2 \text{ and } b_1 > a > b_2. \quad (2)$$

The role of this assumption is further discussed in section 5.

---

[2] Questions related to sequential investigation of the projects, such as the ordering of exploration (cf. Weitzman 1979) should also offer interesting structural results, but I do not address them in this work.

To determine when it makes sense to invest in research in both alternatives, I look at the expressions for the value of information on the two uncertainties jointly and compare that to the value of information on each uncertainty separately. From assumption (2) the value of information on each uncertainty separately is positive, and therefore the joint value of information will also be positive.

Note that I assume perfect resolution of the uncertainties, meaning that once an uncertainty is resolved, the value of the outcome is known for sure. I discuss this assumption in section 5.

First look at the value of information on each alternative separately and then at the value of information for both alternatives jointly.[3] The value of information on alternative A alone is

$$p \max(a_1, b) + (1\text{-}p) \max(a_2, b) - a$$

From assumption (2), with perfect information on alternative A, if you learn that the outcome will be a_1, you prefer a_1 to b; if you learn a_2, you prefer b to a_2, so this expression changes to

$$p\, a_1 + (1\text{-}p)\, b - a \qquad (3)$$

The value of information on alternative B alone is

$$q \max(b_1, a) + (1\text{-}q) \max (b_2, a) - a$$

From assumption (2), if you learn b_1, you prefer b_1 to a; if you learn b_2, you prefer a to b_2, so this expression changes to

$$q\, b_1 + (1\text{-}q)\, a - a \qquad (4)$$

The value of information on both alternatives jointly, assuming they are probabilistically independent is

$$pq \max (a_1, b_1, 0) + p(1\text{-}q) \max (a_1, b_2, 0) + (1\text{-}p)q \max(a_2, b_1, 0) + (1\text{-}p)(1\text{-}q) \max (a_2, b_2, 0) - a \qquad (5)$$

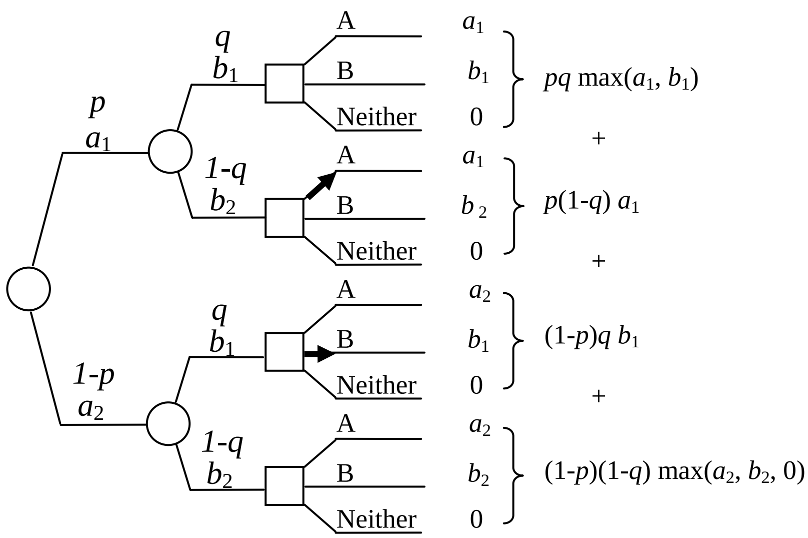
From (1) and assumption (2), I conclude that $a_1 > b_2$ (because $a_1 > a > b > b_2$) and $b_1 > a_2$ (because $b_1 > a > a_2$). So (5) reduces to

$$pq \max (a_1, b_1) + p(1\text{-}q)\, a_1 + (1\text{-}p)q\, b_1 + (1\text{-}p)(1\text{-}q) \max (a_2, b_2, 0) - a \qquad (6)$$

[3] I refer to the value of information on a single uncertainty as "individual information" and to the value of information on both uncertainties together as "joint information."

Figure 1 shows the breakdown of the first four terms of (6). The fifth term is the value of the default alternative (A, because of the assumption $a > b > 0$), the alternative picked in the absence of uncertainty resolution.

Figure 1: The decision tree with joint information.



I assume that the costs of information collection are strictly positive and additive: that is, it costs c_A to collect information on alternative A, c_B to collect information on alternative B, and $c_A + c_B$ to collect information on both. The main conclusion about the value of joint information will be even stronger if the joint cost is less than $c_A + c_B$.

Joint information collection is optimal if it exceeds the value of individual information for each uncertainty by more than the cost of resolving the other uncertainty. In my notation, the joint approach (“diversification”) is optimal when two conditions hold, (6) > (3) and (6) > (4), adjusting for the costs of information collection:

$$pq \max(a_1, b_1) + p(1\text{-}q)\, a_1 + (1\text{-}p)q\, b_1 + (1\text{-}p)(1\text{-}q) \max(a_2, b_2, 0) - a - c_A - c_B > p\, a_1 + (1\text{-}p)\, b - a - c_A \quad (7)$$

and

$$pq \max(a_1, b_1) + p(1\text{-}q)\, a_1 + (1\text{-}p)q\, b_1 + (1\text{-}p)(1\text{-}q) \max(a_2, b_2, 0) - a - c_A - c_B > q\, b_1 + (1\text{-}q)\, a - a - c_B \quad (8)$$

Inequalities (7) and (8) reduce to

$$pq \max(a_1, b_1) + (1\text{-}p)(1\text{-}q) \max(a_2, b_2, 0) > pq\, a_1 + (1\text{-}p)(1\text{-}q)\, b_2 + c_B \qquad (9)$$

and

$$pq \max(a_1, b_1) + (1-p)(1-q) \max(a_2, b_2, 0) > pq\, b_1 + (1-q)(1-p)\, a_2 + c_A \qquad (10)$$

Inequality (9) is satisfied only if $a_1 < b_1$ or $a_2 > b_2$ or $0 > b_2$. Inequality (10) is satisfied only if $a_1 > b_1$ or $a_2 < b_2$ or $0 > a_2$. With three conditions for each inequality, there are nine combinations of conditions. However, two of the combinations are self-contradictory—the combinations of 1) $a_1 < b_1$ and $a_1 > b_1$ and 2) $a_2 > b_2$ and $a_2 < b_2$. By focusing on the relationship between a_1 and b_1, the seven remaining combinations drop down to four, partitioning the possibilities into those with $a_1 < b_1$ and those with $a_1 > b_1$. The resulting four scenarios are as follows.

1. $a_1<b_1$ and $a_2<b_2$
2. $a_1<b_1$ and $0>a_2$
3. $a_1>b_1$ and $a_2>b_2$
4. $a_1>b_1$ and $0>b_2$.

The union of the four scenarios is a necessary condition for simultaneously satisfying both (9) and (10): that is, for joint information to be the best option, at least one of these scenarios must hold. The scenarios are not sufficient, however, because the slack in the inequality must be large enough to overcome the cost of gathering information (c_A and c_B).

In scenarios 1 and 3, the outcomes of one of the alternatives "dominate" the outcomes of the other. For example, in scenario 1, both the high outcome and the low outcome of alternative B are higher than the high outcome and the low outcome of alternative A, respectively.[4] The joint information is valuable in these dominance scenarios because it allows for finer distinctions than in the single information cases. If you gather information only on alternative B, then a low value for B means that you should choose A. But if B "outcomes-dominates" A, and you gather information on both, then even a low value for B may lead to a choice of B in the case that you also see the low value for A.

In scenarios 2 and 4 respectively, alternatives A and B are "outcomes-dominated," not by the other alternative, but by a combination of the other alternative on the high end and the do-

[4] Note that I am using the word dominate here in the narrow sense of outcomes only, and not in the sense of probabilistic dominance. I refer to this relation as "outcomes-dominance."

nothing alternative (i.e., abandoning both) on the low end. Because both alternatives are assumed to be positive in expected value, the abandonment option is never an optimal course of action when gathering information on a single uncertainty. Because joint information adds this new possibility, it can increase the value over single information.

To further explain this logic, look at cases that do not satisfy any of the four scenarios, such as $a_1 > b_1 > b_2 > 0 > a_2$ and $a_1 > b_1 > b_2 > a_2 > 0$. In these cases, joint information will not be any more valuable than information on alternative A alone. Even with joint information, a high value for A will dictate a choice of A; a low value for A will dictate a choice of B.

4. Probabilistic dependence

The preceding analysis assumed probabilistic independence. But what if the outcomes of the two uncertainties are not independent? If the two alternatives address the same market need—e.g., two products to solve the problem of oral care while on the go—market research revealing that one will be popular is likely to reveal the same for the other, implying a positive relationship in their outcomes. However, the outcomes for two alternatives with some significant differences may be negatively correlated—e.g., with technical standards such as those will cellular telephony—the popularity of one standard might imply the likely unpopularity of the other. Either possibility is plausible.

In this section, I will demonstrate that the incremental value of joint information over either piece of individual information is highest with a mild positive dependence between the uncertainties. There are three, somewhat conflicting, forces at play. First, as the dependence becomes extreme in either the positive or negative direction, resolving a single uncertainty provides a strong signal about the other; this influence increases the value of individual information. Second, higher dependence raises the chance that both alternatives will have low outcomes; this influence decreases the value of joint information. Third, higher dependence raises the chance of an intricate comparison, a comparison between two high outcomes or two low outcomes; this influence increases the value of joint relative to individual information.

To construct this argument, first I describe a model of probabilistic dependence, building on the model in section 3. Then I look at the value of information on a single uncertainty to explain the first force, the value of joint information to explain the second, and the comparison between them to explain the third. The comparison also reveals that the four scenarios from section 3 still apply and that within those scenarios, if each alternative has non-zero value of information individually, then the advantage of joint over individual information is greatest with mild positive dependence.

4.1 Model of probabilistic dependence

With dependence, the probability of the joint outcome (a_1, b_1) is no longer the product of the marginal probabilities, pq. With positive dependence, it would be greater than pq, and with negative dependence it would be less than pq. For the best comparison with the case of independent outcomes, I keep the marginal probabilities for the high outcomes for the two alternatives at p and q. I introduce Δ as a measure of the strength of dependence between the outcomes for the two alternatives, with $\Delta \equiv \text{Prob}(a_1, b_1) - pq$. Keeping the marginal probabilities at p and q results in the new probabilities in column 3 of Table 1.

Table 1: New notation for the probabilities of joint events without the independence assumption.

Outcome	Probability under independence	Probability under dependence
(a_1, b_1)	pq	$pq + \Delta$
(a_1, b_2)	$p(1-q)$	$p(1-q) - \Delta$
(a_2, b_1)	$(1-p)q$	$(1-p)q - \Delta$
(a_2, b_2)	$(1-p)(1-q)$	$(1-p)(1-q) + \Delta$

Although Δ is not equal to the correlation coefficient ρ, it is conceptually similar, and in fact there is a linear relationship between Δ and ρ. See Appendix A.1 for the exact form of this relationship. One advantage of using Δ over manipulating ρ is that Δ isolates the effect of dependence, leaving p and q and the outcomes (a_1, a_2, b_1, and b_2) unchanged. One drawback to using Δ to capture dependence is that the range over which Δ can vary is limited. The Δ is constrained to lie within

$$[-\min(pq, (1\text{-}p)(1\text{-}q)),\ \min(p(1\text{-}q), q(1\text{-}p))] \quad (11)$$

to keep all the probabilities non-negative.

4.2 Value of information on a single uncertainty is convex and piecewise linear in Δ, the measure of dependence

Even with the marginal probabilities held at p and q, the value of information on a single uncertainty changes with dependence, because learning which outcome will occur for one alternative affects the estimated value for the other alternative. For example, with positive dependence between the two alternatives (i.e., positive Δ), knowing that a_1 occurred increases the chance of b_1 and knowing that a_2 occurred increases the chance of b_2.

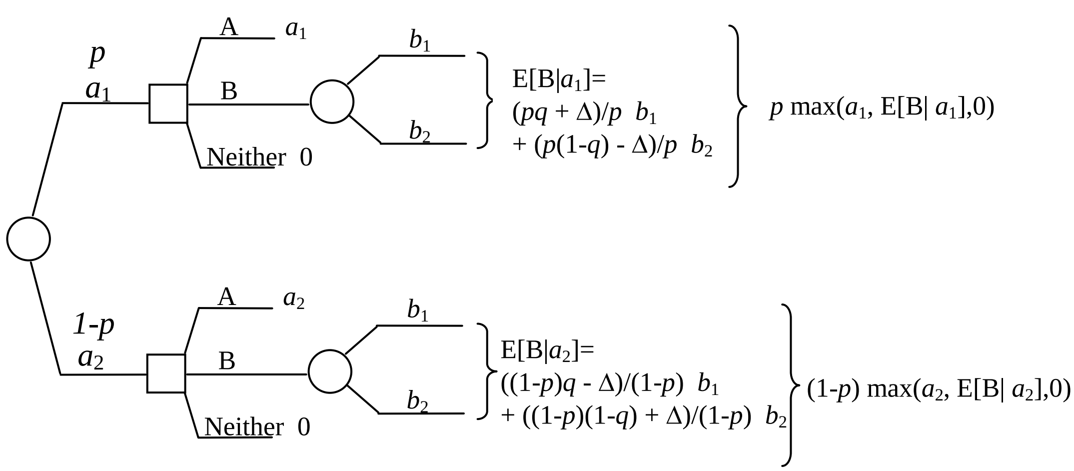
The expressions below, (12) and (13), represent the value of information on A and B, respectively, with probabilistic dependence.

$$\max(pa_1, (pq + \Delta)\, b_1 + (p(1\text{-}q) - \Delta)\, b_2) + \max((1\text{-}p)a_2, 0, ((1\text{-}p)q - \Delta)\, b_1 + ((1\text{-}p)(1\text{-}q) + \Delta)\, b_2) - a \quad (12)$$

$$\max(qb_1, (pq + \Delta)\, a_1 + ((1\text{-}p)q - \Delta)\, a_2) + \max((1\text{-}q)b_2, 0, (p(1\text{-}q) - \Delta)\, a_1 + ((1\text{-}p)(1\text{-}q) + \Delta)\, a_2) - a \quad (13)$$

Figure 2 shows the a decision tree with the logic underlying (12).

Figure 2: The decision tree with information on A and probabilistic dependence.

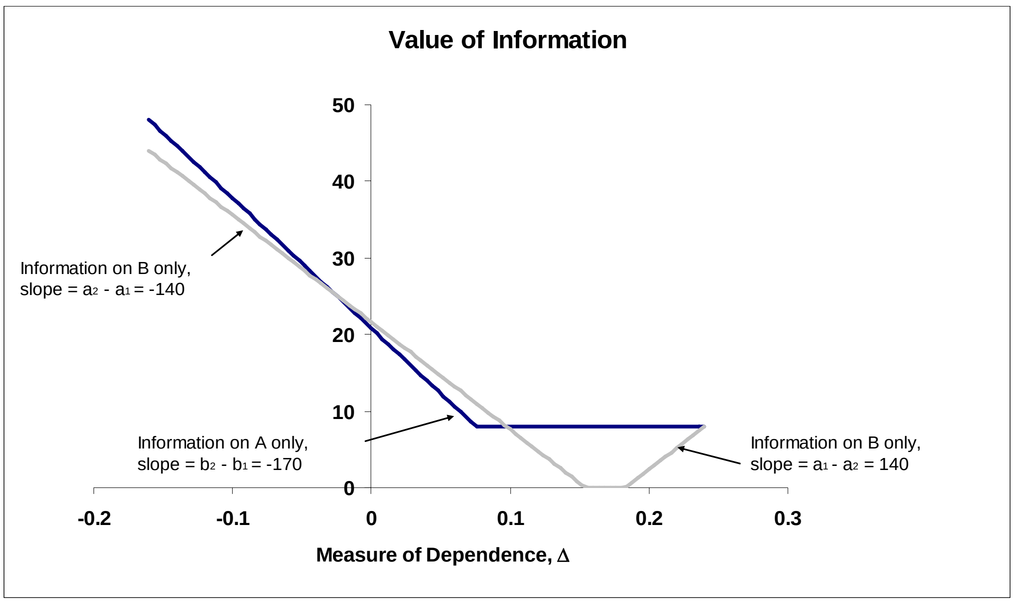


To illustrate the properties of expressions (12) and (13), I introduce a numerical example with

Alternative A	$p = .6$	$a_1 = 120$	$a_2 = -20$	⇒ Mean = 64, and
Alternative B	$q = .6$	$b_1 = 100$	$b_2 = -70$	⇒ Mean = 32.

This example is illustrated in Figure 3.

Figure 3: Value of information of each alternative separately, as a function of the measure of dependence Δ. The range of Δ is limited to that given in (11), which is [-.16, .24] in this numerical example.



The trajectories in (12) and (13) are piecewise linear. The kinks come at Δs that are switchpoints in the maximum functions. For example, the first term in expression (13) is max(qb_1, $(pq + \Delta)\, a_1 + ((1\text{-}p)q - \Delta)\, a_2$). The plot of that term will change slope at the Δ such that $qb_1 = (pq + \Delta)\, a_1 + ((1\text{-}p)q - \Delta)\, a_2$, which is $\Delta = q\, (b_1\text{-}a)/(a_1\text{-}a_2)$. In the numerical example, that value is the lower of the two kinks in the B trajectory, at 0.154. In general, within the Δ range (11), each trajectory can have up to two kinks. The slopes of the lines between the kinks have a symmetric pattern: for A, the slopes (in order) are $b_2\text{-}b_1$, 0, and $b_1\text{-}b_2$; for B they are $a_2\text{-}a_1$, 0, and $a_1\text{-}a_2$. See Figure 3. Appendix A.2 contains the expressions for the switchpoints and the trajectories.

This "tub" shape comes from the fact that when the dependence is strong, either in the positive or negative direction, investing to resolve one uncertainty offers a strong signal about the other. This buy-one-learn-about-the-other logic indicates that with extreme dependence, joint information will not be significantly greater than individual information; joint information will have the greatest advantage somewhere in the middle of the Δ range.

4.3 Value of joint information is decreasing linearly in Δ, *the measure of dependence.*

The value of joint information under independence was given in (6). Given dependence as summarized by the model in Table 1, (1), and (2), the value of joint information is

$$(pq + \Delta) \max (a_1, b_1) + (p(1\text{-}q) - \Delta)\, a_1 + ((1\text{-}p)q - \Delta)\, b_1 + ((1\text{-}p)(1\text{-}q) + \Delta) \max (a_2, b_2, 0) - a \qquad (14)$$

To see that (14) is decreasing in Δ, examine two cases: $a_1 > b_1$ and $a_1 < b_1$.

If $a_1 > b_1$, (14) simplifies to

$$pa_1 + (1\text{-}p)q\, b_1 + (1\text{-}p)(1\text{-}q) \max (a_2, b_2, 0) + \Delta\, [\max (a_2, b_2, 0) - b_1] - a \qquad (15)$$

Because $[\max (a_2, b_2, 0) - b_1]$ is negative, (15) is decreasing in Δ.

If $a_1 < b_1$, (14) simplifies to

$$qb_1 + p(1\text{-}q)\, a_1 + (1\text{-}p)(1\text{-}q) \max (a_2, b_2, 0) + \Delta[\max (a_2, b_2, 0) - a_1] - a \qquad (16)$$

Because $[\max (a_2, b_2, 0) - a_1]$ is negative, (16) is decreasing in Δ.

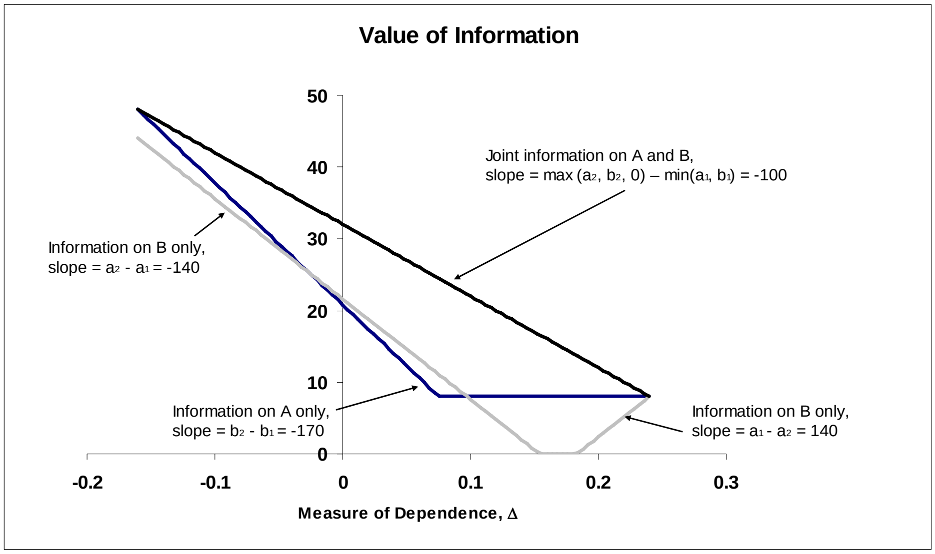
Thus, the value of joint information is a decreasing linear function of Δ with slope of

$$\max (a_2, b_2, 0) - \min(a_1, b_1).$$

As Δ increases, the benefit from the increased chance of two high outcomes does not compensate for the loss from the increased chance of two low outcomes, so the slope of the value of joint information with respect to Δ is negative. The asymmetry comes from being able to select the better of the two alternatives in the final stage, so that getting two high outcomes doesn't necessarily leave you better off than a single high outcome, but two low outcomes are definitely worse than a single low outcome.

Continuing the numerical example from part 4.2, Figure 4 shows the value of joint information in addition to the individual information as a function of Δ.

Figure 4: Value of information of the alternatives both separately and jointly



4.4 Comparing joint and single information

In section 3, I derived the necessary conditions under which joint information was strictly more valuable than both pieces of individual information when the uncertainties are independent. Now, I examine the same question for the model with probabilistic dependence.

We know from Blackwell (1953) that costless joint information cannot have less value than costless individual information, but what are the conditions such that joint information is *strictly* greater than both pieces of individual information? The conditions come from a comparison of the slopes of the trajectories of joint and individual information. If the slope of joint information is strictly greater (in this case, less negative, or shallower) than the slope of the first leg of the single information trajectories, then there will be a gap between the value of joint information and each value of single information. The other two legs of the single information trajectories do not need to be compared to the slope of the joint trajectory, because the former have zero and positive slope, so they can't possibly coincide with the latter. Even with different slopes, the joint and individual trajectories may coincide at the boundaries of the Δ range, which is a byproduct of the buy-one-learn-about-the-other property as dependence moves to the extremes (discussed in section 4.2).

The conditions for the strict inequalities of the slopes follow.

Joint > A only: $\max(a_2, b_2, 0) - \min(a_1, b_1) > b_2 - b_1$ (17)

Joint > B only: $\max(a_2, b_2, 0) - \min(a_1, b_1) > a_2 - a_1$ (18)

Continuing with assumption (2), there are four possible scenarios under which these inequalities hold. They are the same four scenarios under which joint information is strictly greater than individual information under independence from section 3, in which one of the alternatives is either "outcomes-dominant" or "outcomes-dominated." (See Appendix A.3 for the details.) As before, the four scenarios are necessary but not sufficient. Sufficiency comes from a disparity (between sides of the inequality in the four scenarios), great enough to cover the incremental cost of information collection, c_A or c_B.

The ordering of these slopes given in (17) and (18), with the joint information slope less negative than either individual information slope, shows the differential impact of changes in Δ. The value of joint information is not as severely affected by increases in Δ because of the opportunity to make a more intricate comparison.

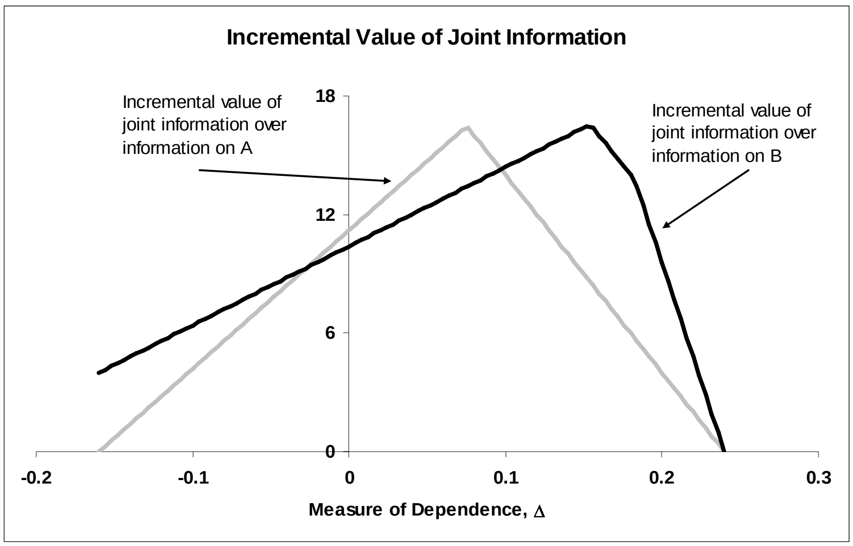
Although it is not transparent how the three conflicting forces—strong positive or negative dependence encouraging buy-one-learn-about-the-other, higher dependence stimulating two low outcomes, and higher dependence facilitating intricate comparisons—will settle up, the preceding analysis supports the following conclusions:

- The greatest difference between the joint trajectory and each single trajectory will be at the first kink in the single trajectory: both trajectories are decreasing, the single one is decreasing faster, and then it flattens or rises at the kinks.
- Assumption (2), specifically $b_1 > a$, implies that the Δ at this maximum difference is at a positive value. (See the Appendix A.4 for the derivation.) In other words, the incremental value of joint information is maximized at a positive value of Δ. Figure 5 illustrates this result for the numerical example. This positive maximum indicates that up to a point, the

value of intricate comparisons outweighs the cost of choosing between two low outcomes.

This second conclusion highlights just how different the risk aversion rationale for diversification is from my information gathering rationale: in the former, extreme negative dependence maximizes the attractiveness of diversification; in the latter, mild positive dependence does.

Figure 5: The highest incremental value of joint information occurs at a positive Δ.



5. Robustness

Section 3 contained the simplest model possible to describe the rationale of the optimality of joint information collection. In this section, I discuss the robustness of that reasoning to the structural assumptions of perfect information, the two-outcome uncertainties, the two alternatives, a strict either/or decision in the final period, and the ordering assumptions on outcomes.

While the model described above is set out as a model of perfect information resolution on uncertainties that have two outcomes, it can also be thought of as a model of imperfect information resolution on a more general distribution. With imperfect information, the two possible outcomes serve the role of the expected values based on information from a test with a binary result.

The logic of my argument, that investigating multiple avenues simultaneously allows for comparisons that could not be made with the results of a single investigation, is not limited to a two-outcome world. If the outcomes are distributed over a continuous range, then there are more possibilities for joint investigation to yield an intricate comparison, leading to a different decision than individual investigation. With continuous outcomes, as with discrete, the value of joint information is still enhanced by one alternative having both a better high-end and a better low-end outcome than the other alternative. In addition, the conclusion that extreme positive or negative dependence in the uncertainties favors single information over joint information still holds.

Allowing for more than two alternatives does not corrupt the argument. In fact, it strengthens the argument, because single investigations narrow down the choices to the investigated alternative and the best of the remaining alternatives, whereas multiple investigations allow for any of the alternatives to be chosen.

If the second stage of the decision problem (the investment stage that comes after the information collection stage) is not actually an either/or situation, then my conclusions need some qualification. If either, neither, or both alternatives can be done, and the alternatives are completely independent, both in probabilities and in values, then contrary to the conclusions in sections 3 and 4, the decisions about the two alternatives decouple. In other words, in that case, the investigation and investment of each alternative can be considered separately from the other alternative; mathematically, the value of information on the two uncertainties is additive. This complete independence is contrary to the setting of this paper, that of evaluation of "competing alternatives." The alternatives can be competing either for consumer attention or firm resources, or both. If both alternatives can be done, but the value of each alternative is reduced when they are both done (e.g., total sales of a product line would be higher than one product alone, but each product in the line would sell more if it were the only product), the interdependence of the investigative investments still holds.

In sections 3 and 4, the calculations were done under assumption (2). This assumption ruled out the possibility of deterministic dominance of A over B: that is,

$$a_1 > a_2 > b_1 > b_2.$$

In this case there is no value of information either singly or jointly because no matter how the uncertainty turns out, A is always better than B.

Further, assumption (2) is important in another way. Assumption (2) posits that both alternatives have non-zero value of information individually. If both alternatives have zero value of information individually under independence, i.e., $a_2 > b$ and $a > b_1$, then the first switchpoint (from Figure 3 and Appendix A.2) for each occurs at a negative value. So if both have zero value of information separately, the incremental value of joint information over individual information is maximized at a mildly negative Δ. In this case, mild negative dependence, with its increased chance of at least one alternative's having the good outcome, results in higher incremental joint value than positive dependence, with its increased chance of intricate comparisons.

6. Discussion

The former student with the network security company mentioned in the introduction faced a fork in the road. The company started with support from friends and family funding, but needed to show beta-level software in trials at client companies to attract to larger-scale investors. At the time of his decision, the wired market was bigger and more mature, and the wireless market was smaller and growing. With rough estimates of the addressable market size in both markets, but no substantial market research due to financial and time constraints, the company chose to go after the small but growing wireless security market. While wireless networking did, in fact, take off, their software-based approach fell out of favor as network equipment makers incorporated more security features directly into the network hardware.

With a bit less pressure on time and money, the firm may have been tempted to do more market research on the wireless opportunity (in their mind, the front runner) to confirm its superiority or on the wired opportunity (the underdog) to rule it out conclusively. One might think that confirming the better alternative is the way to go when that alternative has a higher mean and dominates (either in the weak sense used above or in the usual definition of stochastic dominance). But I have shown that the dominating case is actually one of the scenarios in which it is most promising to investigate both alternatives.

When the alternatives are not probabilistically independent, perhaps it is more intuitive that there may be synergy in investigating them jointly compared to independent alternatives. However, extremes of dependence reduce the relative advantage of joint investigation. When both alternatives have non-zero value of information separately, it is mild positive dependence that gives the biggest boost to joint information collection. When two alternatives aim to address the same consumer need, there is likely to be positive dependence in their outcomes. Investing in one only holds the following risk: find out the low outcome, pick the other, but because of the positive dependence, the other is bad too. For the network security company, the underlying growth of the need for all types of networking is a factor that positively impacts both markets, suggesting positive dependence.

I have not emphasized the cost of information collection, but it plays a starring role in the argument. In the extreme, if there is no cost of information collection, then it is better to have both pieces of information. If the information cost is low enough (even if the information is imperfect), having both pieces of information will be better than having only a single piece of information. This paper’s argument is perhaps most interesting when the costs of information collection are such that they would prohibit investing for either alternative in isolation, but promote investing in both alternatives jointly. For the network security company, cost and reliability of market research were two real obstacles to helpful information collection.

I conclude with two additional examples. In 2006, we are witnessing an encore performance of the VCR format wars of the early 1980s, with the two contenders for the next generation DVD standard, Sony’s Blu-Ray and Toshiba’s HD-DVD (Belson 2006). Consumer electronics firms face a dilemma about which of these standards to support. With this type of uncertainty, the traditional market research techniques or technical feasibility studies are of limited help. The standards uncertainty will be resolved only gradually, as the market evolves. The way to think about investment in uncertainty resolution here is as a cost of delay. Admittedly, the cost associated with delay is uncertain (unlike my model, with known cost of uncertainty resolution). However, the logic of my conclusions still applies: With a strong negative correlation between the successes of these two standards, waiting to learn the fate of one will be enough information.

A second example relies more on the traditional type of investment in research. In the pharmaceutical industry, large firms maintain abundant new product development portfolios. GlaxoSmithKline's February 2006 Product Development Pipeline (GSK, 2006) shows three compounds in Phase I and II trials for the indication of dyslipidaemia, three for osteoporosis, four for rheumatoid arthritis, and nine for type 2 diabetes. In this context, certainly more than one reason for diversification applies: perhaps risk aversion plays a role, and product line considerations are important, with some delivery mechanisms (e.g., injection, pill) more suitable for some patients. But even if those other reasons apply, the power of joint information gathering for intricate comparisons also supports investing in competing approaches. If the probabilities of technical success are independent, and the probabilities of market success are positively correlated, then the overall correlation will be mildly positive within an indication, and that is exactly the situation in which joint information collection is likely to have the most value.

References

Aaker, David A., V. Kumar, and George S. Day. (2004) *Marketing Research*, Eighth Edition. John Wiley & Sons, Inc., New York.

Anderton, E. J., K. Gorton, and R. Tudor. (1980) The Application of Sequential Analysis in Market Research. *Journal of Marketing Research*, 17(1):97-105.

Bain, J.S. (1956) *Barriers to New Competition*. Harvard University Press, Cambridge.

Belson, Ken. (2006) In Sony's Stumble, the Ghost of Betamax. *The New York Times*. February 26, 2006.

Blackwell, D. (1953) Equivalent Comparisons of Experiments. *Annals of Mathematical Statistics* 24(2): 265-272.

Chatterjee, Rabikar, Jehoshua Eliashberg, Hubert Gatignon and Leonard M. Lodish. (1988) A Practical Bayesian Approach to Selection of Optimal Market Testing Strategies. *Journal of Marketing Research*, 25(4):363-375.

Cohen, Morris A., Jehoshua Eliashberg, and Teck H. Ho. (1997) An Anatomy of a Decision-Support System for Developing and Launching Line Extensions. *Journal of Marketing Research*, 34(1):117-129.

Cohen, Wesley M. and Daniel A. Levinthal. (1994) Fortune favors the prepared firm. *Management Science*, 40(2):227–251.

Dixit, Avinash. (1979) A Model of Duopoly Suggesting a Theory of Entry Barriers. *The Bell Journal of Economics*, 10(1):20-32.

Delquié, Ph. (2005) Value of Information vs. Intensity of Preference: A General Result. Working Paper, INSEAD, Fontainebleau.

Dickinson, Michael W., Anna C. Thornton, and Stephen Graves. (2001) “Technology portfolio management: Optimizing interdependent projects over multiple time periods.” *IEEE Transactions on Engineering Management*, 48(4):518–527.

Farley, John U., Donald R. Lehmann, and Lane H. Mann. (1998) “Designing the Next Study for Maximum Impact." *Journal of Marketing Research*, 35 (November): 496-501.

GlaxoSmithKline Product Development Pipeline. (2006) http://www.gsk.com/investors/product_pipeline/docs/pipeline.pdf, accessed March 10, 2006.

Hilton, R.W. (1981) The determinants of information system value: Synthesizing some general results. *Management Science* 27(1): 57-64.

Loch, Christoph H. and Stylianos Kavadias. (2002) Dynamic portfolio selection of NPD programs using marginal returns. *Management Science*, 48(10):1227–1241.

Malhotra, Naresh K. and Mark Peterson. (2006) *Basic Marketing Research: A Decision-Making Approach*, Second Edition. Pearson Education, Inc., Upper Saddle River, New Jersey.

Markowitz, Harry M. (1959) *Portfolio Selection: Efficient Diversification of Investment*. Cowles Foundation, New Haven, Connecticut.

Mehrez, Abraham. (1989) Comments on the willingness of an economic agency to pay for parallel research efforts. *The Journal of the Operational Research Society*, 40(8):787–788.

Mehrez, Abraham and Awanti P. Sethi. (1989) “Hierarchical planning of project-selection problems with information purchasing.” *The Journal of the Operational Research Society*, 40(3):267–279.

Moorthy, K. Sridhar. (1984) Market Segmentation, Self-Selection, and Product Line Design. *Marketing Science*, 3(5):288-307.

Prastacos, Gregory P. (1983) Optimal sequential investment decisions under conditions of uncertainty. *Management Science*, 29(1):118–134.

Pringle, David. (2004) Wrong Number: How Nokia Chased Top End of Market, Got Hit in Middle; Rivals' Cheaper, Stylish Models Gobble Up Market Share As Fancier Ones Languish; Missing Clamshell Signals. *The Wall Street Journal*. June 1, 2004. A1.

Pratt, John W., Howard Raiffa, and Robert Schlaifer. (1995) *Introduction to Statistical Decision Theory*. The M.I.T. Press, Cambridge, Massachusetts.

Raiffa, Howard and Robert Schlaifer. (1961) *Applied Statistical Decision Theory*. The M.I.T. Press, Cambridge, Massachusetts.

Teece, David J. (1980) Economies of Scope and the Scope of the Enterprise. *Journal of Economic Behavior and Organization*, 1(3):223-247.

Teece, David J. (1982) Towards an Economic Theory of the Multiproduct Firm. *Journal of Economic Behavior and Organization*, 3(1):39-63.

Urban, Glen L. and Gerald M. Katz. (1983) “Pre-Test-Market Models: Validation and Managerial Implications.” *Journal of Marketing Research*, 20(3):221-234.

Weitzman, Martin L. (1979) Optimal search for the best alternative. *Econometrica*, 47(3):641–654.

Zaltman, Gerald. (1997) Rethinking market research: Putting people back in. *Journal of Marketing Research* 34(4):424-437.

Appendix

A.1 Relationship between correlation coefficient ρ and Δ:

$$\rho = \frac{\mathrm{cov}(A,B)}{\sigma_A \sigma_B} = \frac{(pq+\Delta)a_1 b_1 + (p(1-q)-\Delta)a_1 b_2 + ((1-p)q-\Delta)a_2 b_1 + ((1-p)(1-q)+\Delta)a_2 b_2 - ab}{\sqrt{pa_1^{\ 2} + (1-p)a_2^{\ 2} - a^2}\sqrt{qb_1^{\ 2} + (1-q)b_2^{\ 2} - b^2}}$$

$$= \frac{pqa_1 b_1 + p(1-q)a_1 b_2 + (1-p)qa_2 b_1 + (1-p)(1-q)a_2 b_2 - ab + \Delta(a_1 - a_2)(b_1 - b_2)}{\sqrt{pa_1^{\ 2} + (1-p)a_2^{\ 2} - a^2}\sqrt{qb_1^{\ 2} + (1-q)b_2^{\ 2} - b^2}}$$

A. 2 Details for the trajectories for the value of individual information for alternatives A and B. The tables below show the value of information functions that apply to the segments of the Δ axis.

Alternative A

Condition on Δ	Value of information on A
$\Delta < (1-p)\ (b-\max(a_2,0))/(b_1 - b_2)$	$pa_1 + ((1-p)q - \Delta)\ b_1 + ((1-p)(1-q) + \Delta)\ b_2 - a$
$(1-p)\ (b-\max(a_2,0))/(b_1 - b_2) < \Delta < p\ (a_1-b)/(b_1-b_2)$	$pa_1 + \max((1-p)a_2, 0) - a$
$\Delta > p\ (a_1-b)/(b_1-b_2)$	$(pq + \Delta)\ b_1 + (p(1-q) - \Delta)\ b_2 + (1-p)\max(a_2,0) - a$

The order of the segments in the table above is fixed because the two switchpoints are in order: $p\ (a_1-b)/(b_1-b_2) > (1-p)\ (b-\max(a_2,0))/(b_1 - b_2)$), which relies on the convention throughout this paper that $a > b$. For the case $a_2 < 0$, it follows from that convention that $pa_1 > b$.

Alternative B

If $qb_1 < a$

Condition on Δ	Value of information on B
$\Delta < q\ (b_1-a)/(a_1-a_2)$	$qb_1 + (p(1-q) - \Delta)a_1 + ((1-p)(1-q) + \Delta)\ a_2 - a$
$q\ (b_1-a)/(a_1-a_2) < \Delta < (1-q)\ (a-\max(b_2,0))/(a_1 - a_2)$	0
$\Delta > (1-q)\ (a-\max(b_2,0))/(a_1 - a_2)$	$(pq + \Delta)\ a_1 + ((1-p)q - \Delta)\ a_2 + (1-q)\max(b_2, 0) - a$

If $qb_1 > a$ ($\Rightarrow b_2 < 0$)

Condition on Δ	Value of information on B
$\Delta < (1\text{-}q)\, a/(a_1 - a_2)$	$qb_1 + (p(1\text{-}q) - \Delta)a_1 + ((1\text{-}p)(1\text{-}q) + \Delta)\, a_2$
$(1\text{-}q)\, a/(a_1 - a_2) < \Delta$ $< q\,(b_1\text{-}a)/(a_1\text{-}a_2)$	$qb_1 - a$
$\Delta > q\,(b_1\text{-}a)/(a_1\text{-}a_2)$	$(pq + \Delta)\, a_1 + ((1\text{-}p)q - \Delta)\, a_2 - a$

A.3 The scenarios for positive incremental value of joint information in the dependence case are the same four scenarios as in the independence case.

If $a_1 < b_1$, then 1) comparing the slope of joint to the slope of A, max $(a_2, b_2, 0) > b_2 + (a_1\text{-}b_1)$ is always true, and 2) comparing the slope of joint to the slope of B, max $(a_2, b_2, 0) > a_2$ is true if $b_2 > a_2$ or $0 > a_2$.

This results in two scenarios under which the inequalities strictly hold:

$a_1 < b_1$ and $b_2 > a_2$

$a_1 < b_1$ and $0 > a_2$

If $a_1 > b_1$, then (1) comparing the slope of joint to the slope of A, max $(a_2, b_2, 0) - b_1 > b_2 - b_1$ is true if $b_2 < a_2$ or $0 > b_2$, and (2) comparing the slope of joint to the slope of B, max $(a_2, b_2, 0) - b_1 > a_2 - a_1$ is always true.

This results in two scenarios under which the inequalities strictly hold:

$a_1 > b_1$ and $b_2 < a_2$

$a_1 > b_1$ and $0 > b_2$

A.4 The switchpoints in the value of individual information trajectories are all positive.

The first switchpoint for the A line will is $(1\text{-}p)\,(b\text{-}\max(a_2,0))/(b_1 - b_2)$. It is positive by assumption (2), with $b > a_2$.

The first switchpoint for the B line could be either $(1-q)\ a/(a_1 - a_2)$, which is always positive, or $q\ (b_1-a)/(a_1-a_2)$, which is positive by assumption (2), with $b_1 > a$.

Assumption (2) insures that the value of individual information in the independent case is positive.