Are Information-gathering and Producing Complements or Substitutes?

Thomas Marschak* J. George Shanthikumar† Junjie Zhou‡

December, 2013

*Walter A. Haas School of Business, University of California, Berkeley. e-mail: marschak@haas.berkeley.edu
†Krannert School of Management, Purdue University. e-mail: shanthikumar@purdue.edu
‡Shanghai University of Finance and Economics. e-mail: zhoujj03001@gmail.com

Abstract

We aim at some simple theoretical underpinnings for the study of a complex empirical question studied by labor economists and others: does Information-technology improvement lead to occupational shifts — toward "information workers" and away from other occupations — and to changes in the productivity of non-information workers? In our simple model there is a Producer, whose payoff depends on a production quantity and an unknown state of the world, and an Information-gatherer (IG) who expends effort to learn more about the unknown state. The IG's effort yields a signal which is conveyed to the Producer. The Producer uses the signal to revise prior beliefs about the state and uses the posterior to make an expected-payoff-maximizing quantity choice. Our central aim is to find conditions on the IG and the Producer under which more IG effort leads to a larger average production quantity (Complements) and conditions under which it leads to a smaller average quantity (Substitutes). For each of the IG's possible efforts there is an information structure, which specifies a signal distribution for every state and (for a given prior) a posterior state distribution for every signal. We start by considering a Blackwell IG. For such an IG, the possible structures can be ranked so that a higher-ranking structure is more useful to every Producer, no matter what the prior and the payoff function may be. For the Blackwell IGs whom we consider, a higher-ranking structure is reasonably interpreted as a higher-effort structure. The Blackwell theorems state that one structure ranks above another if and only if the expected value (over the possible signals) of any convex function on the posteriors is not less for the higher-ranked structure. So we have Complements (Substitutes) if the Producer's best quantity is indeed a convex (concave) function of the posteriors. That gives us Complements/Substitutes results for a variety of Producers. We then turn to a non-Blackwell IG who partitions the state set into n equal-probability intervals. The IG can choose any positive integer n and n is the effort measure. We recapture some of the results from the Blackwell-IG case, but far different techniques are needed, since the Blackwell theorems cannot be used.

Keywords: Information technology and productivity, Blackwell Theorem, Garbling,

JEL classifications: C44, D24, L11

Contents

1 Introduction **1**

1.1 Information-technology advances, productivity surges, and occupational shifts. . . 1

1.2 A preview of some results 5

1.3 Related literature. 8

1.4 The formal framework. 9

1.5 Four Blackwell theorems 10

1.6 Plan of the rest of the paper. 17

1.7 Notation and terminology. 19

2 The Blackwell Information-gatherer **19**

2.1 Obtaining Complements/Substitutes results with and without the Blackwell theorems. 19

2.2 Strict garbling of finite structures 20

2.3 Two Blackwell families of finite information structures, where each structure is defined by a parameter $k \in [0, 1]$. 23

2.3.1 A family in which each structure is a linear combination of two "anchor" structures: introduction. 23

2.3.2 Results that hold with no further restrictions on the two anchor experiments. 25

2.3.3 Requiring the anchor structure I_0 to be a null structure. 28

2.3.4 Following a different path to the statements in Theorem B4: calculating a garbling matrix and applying the First Blackwell Theorem. 31

2.3.5 Some remarks on the family of structures $\{I_k\}_{k\in[0,1]}$ considered in the preceding theorems. 32

2.4 Counterparts of the previous results when the state and signal sets are continua. . 35

2.5 The *erratic IG*: another Blackwell family of finite information structures indexed by $k \in [0, 1]$, where k is the probability that the IG uses the "better" of two structures. 40

2.6 Sufficient conditions for best quantity to be a convex (concave) function of the posterior. 43

3 Non-Blackwell information-gatherers. **44**

3.1 Three theorems about an IG who constructs equal-probability partitionings, or equal-length partitionings, of the state set. 44
3.2 Two questions suggested by Theorem N1 47
3.2.1 Is Theorem N1 implied by basic propositions concerning stochastic orders? 47
3.2.2 Do the conclusions of Theorem N1 hold for a much wider class of distributions on $[0, 1]$? 48
3.3 A non-Blackwell Information Gatherer who performs a scale/location transform on a base distribution. 48

4 Some Complements/Substitutes results that are implied by the preceding Theorems. 50
4.1 A Blackwell IG and a Producer whose best quantity depends on the posterior mean: applying the Blackwell Theorems and Theorems B1-B3. 50
4.1.1 A price-taking Producer who is uncertain about price. 52
4.1.2 A monopolist who is uncertain about demand. 53
4.1.3 price-taker who is uncertain about cost. 55
4.1.4 Applying Theorems B1-B4 or Theorems $\tilde{B}1 - \tilde{B}4$ when the IG's structures are linear combinations of two anchor structures. 56
4.2 Applying Theorems N1-N3: a non-Blackwell IG partitions the states into equal-probability or equal-length intervals and conveys signals to a Producer whose best quantity depends on the posterior mean. 56
4.3 An application of the second strict-garbling theorem. 57

5 Concluding remarks. 58

A Proofs 59
A.1 Proof of the first strict-garbling theorem in section 2.2. 59
A.2 Proof of the second strict-garbling theorem in section 2.2. 61
A.3 Proof of Theorem B5 63
A.4 Proof of Theorem $\tilde{B}1$. 63
A.5 Proof of Theorem $\tilde{B}5$. 65
A.6 Proof of Theorem B6. 66
A.7 Proof of Theorem B7. 67
A.8 Proof of Theorem N1 69

A.9 Proof of Theorem N2 76
A.10 Proof of Theorem N3 77

B Constructing the counterexample discussed in Section 3.2.2 79

1 Introduction

1.1 Information-technology advances, productivity surges, and occupational shifts.

This paper aims to contribute some simple theoretical underpinnings to the study of several complex but empirically well-motivated questions: do advances in Information Technology, IT investments, and more effort by “information workers” lead producers to produce more or to produce less, to use more non-information workers or fewer? Do those workforce shifts raise or lower the productivity of the non-information workers the producers employ?

Empirical economists have tried for many years to trace the impact of advances in Information Technology (or, in a more recent terminology, “Information and Communication Technology”) on productivity. The prevailing approach studies a firm, an industry, or a country and assembles time-series data on IT-related capital, IT-using labor, non-IT capital, non-IT labor, and output. A Cobb-Douglas relation between output and inputs is typically assumed and the coefficients for IT capital and IT labor are of particular interest.[1] Many studies conclude that observed differences between US productivity and that of other economies, as well as the US productivity surge of the late 1990s, are indeed well “explained” — in the statistical sense — by changes in IT input quantities and improvements in the technology itself. [2]

A related phenomenon is an occupational shift in the US labor force. The proportion of “blue-collar” workers (notably those with low-skill manufacturing jobs) has declined as well as their relative wage. Has there been a corresponding rise in the proportion of information workers and their relative wage? Defining “information workers” in a useful way is challenging. A key contribution is Autor, Levy, and Murnane (2003). They are interested in the effects of increased computer investment on the hired quantities of two types of labor: labor that performs routine tasks and labor that performs nonroutine cognitive tasks. Here we are interested in workers of the second type, whose main task is to acquire and analyze information about the randomly changing environment in which the employer operates, so as to improve the quality of the employer’s decisions. When we observe more workers of this type being hired, or improved information technology being placed at their disposal, will we also see the employer producing, on the average, a higher quantity or a lower one? Is the change in average production quantity accompanied by increased or decreased hiring of non-information workers? Tracking the kind of workers we have in mind is difficult. A hint of the difficulty is found in O*NET (the Occupational Information Network developed by the Employment and Training Administration of the US Department of Labor). One finds 867 occupations linked to the term ”information processing”.

[1] A recent survey of this literature is T. Kretschmer (2012).

[2] For example, Jorgenson, Ho, and Shiroh (2005) claim with confidence that the dramatic productivity gains of the 1990s are largely explained by IT advances. They compare four IT-producing industries, 17 IT-using industries, and 23 non-IT industries. The IT-producing and IT-using industries account for 30 % of GDP but contributed half of the rise in GDP growth.

[3,4]

It is one thing to track occupation shifts and productivity changes, and statistically estimate the esxplanatory power of certain observable variables. It is quite another to model organizations that use IT inputs and non-IT inputs and to find conditions under which improvements in information technology or a drop in the price of IT inputs leads the organization to shift its workforce in a given direction, perhaps causing a change in the productivity of its non-IT inputs. Thus an observed shift away from low-skill manufacturing workers and towards information workers, might be due to "off-shoring" of manufacturing. But it might also be the case that increasing the organization's information workforce, thereby improving its decisions, leads it to increase (decrease) its average production quantity which leads, in turn, to the hiring of more (fewer) non-information workers.

Several papers have suggested that improvements in IT induce the organization to change its organizational structure in ways that complement its IT investment and enhance the productivity of both IT-using labor and non-IT labor.[5] To explore this complex conjecture in a formal way would require careful modeling of the organization's goals and activities. We shall not attempt a model of organization structure.

Our framework will be much more modest. We study just two idealized persons: an Information-gatherer (an IG) and a Producer. The Producer's payoff depends on the product quantity that he chooses and on a state of the world that is unknown at the time the choice has to be made. The Producer has beliefs about the unknown state, expressed as a prior probability distribution. Before choosing his quantity, the Producer acquires the IG's services. The IG expends effort to learn more about the unknown state and sends a signal to the Producer, who uses the signal to replace the prior distribution with a posterior one. Using the posterior distribution, the Producer chooses a quantity that maximizes expected payoff. At each new level of effort, the IG has a fresh signal set, implying a fresh set of possible posteriors.

We ask: when will increased IG effort lead the Producer to choose, on the average, a higher quantity (which may require more non-informational inputs)? When will it lead the Producer to choose, on the average, a lower quantity? It will be convenient to call the first case *Complements* and to call the second case *Substitutes*.

Consider the following examples. In each of them, the term "more IG effort" is informal. It

[3]A recent extensive discussion of the task identification problem is found in D. Acemoglu and D. Autor, "Skills, Tasks and Technologies: Implications for Employment and Earnings", *Handbook of Labor Economics, Volume 4b*, 2011.

[4]A possible proxy for the analytic information workers we have in mind might be what the Bureau of Labor Statistics calls "STEM" (Science, technology, engineering, and mathematics) occupations. The proportion of workers in these occupations, and their relative wages, have indeed grown rapidly, with far faster growth in some industries than in others. See B. Cover, J. Jones, and A. Watson (2011), M. Lee and M. Mather (2008); D. Stine and C. Matthews (2009).

[5]See Bresnahan, Brynjolffson, and Hitt (2002), and Hubbard (2000, 2003)(on the trucking industry).

might mean more hours or more investment in information-processing capitals, or both.

- The Producer is an expected-profit-maximizing manufacturer who takes product price as given and is uncertain about price, or faces a linear demand curve and is uncertain about one of its parameters. The IG collects and studies appropriate data and makes a point-valued forecast of the unknown price or parameter. More IG efforts makes the forecast more reliable, in a precise sense, as well as more useful to the Producer. When the IG exerts more effort, does the producer choose a higher average quantity (averaging over the possible forecasts) or a lower average quantity?

- The Producer is again an expected-profit-maximizing manufacturer who is uncertain about demand. There is a population of potential demanders. There are two demander types. A Type-A demander is willing to pay up to X dollars for one unit. A Type-B demander is unwilling to buy anything, whatever the price may be. The IG samples the population, learns the sampled demanders' true types, and tells the Producer the proportion of Type-A demanders in the sample. More IG effort means a larger sample. When the IG exerts more effort, does the average product quantity chosen by the Producer rise or fall?

- Again the Producer is an expected-profit-maximizing manufacturer who takes product price as given or faces a linear demand curve. He knows the price or the demand curve, but his total cost for the quantity q is $\theta C(q)$, where θ is a random variable with a known prior distribution. The IG does appropriate research on the determinants of cost and then sends the Producer a signal which leads the Producer to replace the prior by a posterior. More IG effort makes the posterior expected value of θ closer, in a precise sense, to the true θ, and makes the posterior more useful to the Producer. When the IG exerts more effort, does the Producer choose a higher average quantity or a lower one?

- The Producer is an expected-profit-maximizing inventory manager who sells product at a fixed and known price, but has to order before knowing what the next period's demand will be. The IG forecasts demand and more IG effort makes the forecast more reliable, in a precise sense, and more useful to the Producer. When the IG exerts more effort, does the Producer, on the average order more or less?

- The Producer is the inventory manager of the preceding example and the IG is the same as well. The Producer orders optimally, given the IG's forecast. The Producer sells the order or the realized demand, whichever is smaller. When the IG exerts more effort, does the Producer, on the average, sell more or less?

- The Producer is a lender who is uncertain about borrowers' qualifications. The IG probes the qualifications, computes a credit score for each borrower, and sends the score to the Producer. More IG effort means that the reported score is, in an appropriate sense, a more reliable indicator of the borrower's future behavior and is more useful to the Producer. When the IG exerts more effort, does the Producer, on the average, lend more or less?

- The Producer owns a fleet of trucks and the IG is a dispatcher who examines randomly changing traffic on alternate routes and advises each trucker as to the routes that look most promising. The IG's characterization becomes more reliable when he exerts more effort. When routes are used which turn out to be less congested in a given week, more customers receive their deliveries that week for a given number of trucks dispatched that week. When the IG exerts more effort, will the Producer choose to send out more trucks in the average week or fewer?[6]

- Going further afield, the Producer is a surgeon who performs a procedure following a certain diagnostic result. The IG is a diagnostician. By exerting more effort he can reduce the frequency of false positives. When the IG exerts more effort, does the Producer, on the average, perform more or fewer procedures?

Note that in each example more precise modeling of the IG and his efforts is needed. In the last three examples we also need more precise modeling of the Producer's payoff and the way it depends on the randomly changing state of the world.

Our Complements/Substitutes questions concern the *direction* in which the Producer's chosen product quantity moves when the IG works harder and not, in general, the size of that movement or the size of the IG's effort. To study the Producer's "best" choice of IG effort would require balancing the cost of extra effort against the expected-payoff increment that it yields. That is, in general, difficult. Nevertheless we occasionally pose Complements/Substitutes questions in the following way. If the producer balances the benefit and the cost of additional IG effort, and the cost of effort drops, does the Producer choose, on the average, a higher product quantity (Complements) or a lower product quantity (Substitutes)? For one class of IGs [7], effort will be a number in $[0, 1]$. We will let the price of a unit of effort change as effort increases. We will be able to characterize the effort-price functions for which a best effort lies in the interior of $[0, 1]$.

A more thorough study of the Producer's best choice of IG effort would include incentive issues.We defer incentive issues to subsequent research. The deferred research would address the truthfulness of the IG's reports to the Producer. It would recognize that if the IG is free to choose his effort (which may be hidden), bears the effort's cost, and is rewarded by the Producer, then the Producer seeks a reward scheme that is best for him among all the truth-inducing schemes to which the IG will agree. We would then have a new Complements/Substitutes question: if the IG's cost for every effort drops (because technology improves), will there be a rise or a fall in the Producer's average quantity when he responds to the signals that the IG sends him under a best reward scheme?

One can certainly worry as to whether the answer to such a difficult question might reverse the direction of a Complements/Substitutes result that we obtain in our much simpler

[6] This situation is studied from an incentive point of view in Hubbard (2000, 2003).

[7] Discussed in 2.3 below.

incentive-free framework. As we shall see, however, it is quite challenging to obtain general Complements/Substitutes results even in an incentive-free model. Such models are a natural starting place.

The preceding examples suggest that simple intuition is unlikely to help. A very primitive intuition in favor of Complements might say:

> When a Producer learns more about the merits of alternative production decisions, then a dollar devoted to producing earns more, so he will produce more.

A very primitive intuition in favor of Substitutes might say:

> When a Producer learns more about the outside world, then he is better able to protect himself against bad contingencies and does not need as a high a production capacity to achieve a given level of protection. So he will choose a smaller capacity and a smaller average production quantity.

These intuitions are unhelpful. We have to study specific models of the IG and specific models of the Producer. For each model pair, we shall consider precise conditions on the Producer's payoff function, the IG's procedures, and the probabilities of the unknown states-of-the-world. We seek conditions which imply Complements and conditions which imply Substitutes.

1.2 A preview of some results

Consider the first of the preceding examples. The Producer's payoff is

$$u(q, \theta) = \theta q - C(q),$$

where θ is a randomly changing price, $C(q)$ is total cost, and C is thrice differentiable with $C(0) = 0, C' > 0, C'' > 0$. Let the true price θ have n possible values, denoted $\theta_1, \ldots, \theta_n$. The Producer has prior probabilities on the n values. The Producer receives a signal y from the IG and then chooses the nonnegative quantity $q^*(y)$, which maximizes the posterior expected payoff $E(u(q, \theta) \mid y)$ on the set of all nonnegative quantities.

Now suppose the IG is a *one-point forecaster*. (We provide this IG in detail in Remark 2.3.5.2).

Each of the IG's signals is a forecast, identifying one of the n possible prices. The signal y_i forecasts that the true price is θ_i. Given that θ_i is indeed the true price, the forecast y_i is correct with probability x and incorrect with probability $\frac{1-x}{n-1}$, where $\frac{1}{n-1} \leq x \leq 1$.

For the hardest-working IG, x equals 1 and the forecast is always correct. For the least hard-working IG, x is $\frac{1}{n-1}$ and all forecasts are equally probable, so the signal tells the Producer nothing. As the IG works harder, x moves further from $\frac{1}{n-1}$ and closer to 1. It turns out that:

- If $C'''(q) > 0$ at all $q \geq 0$, then we have Substitutes: when the IG's effort x increases, there is a *drop* in the average value of the Producer's best quantity $q^*(y)$ (over all of the n possible signals y).
- If $C'''(q) < 0$ at all $q \geq 0$, then we have Complements: when the IG's effort x increases, there is a *rise* in the average value of the Producer's best quantity $q^*(y)$ (over all of the n possible signals y).
- Regardless of the sign of C''', more IG effort benefits the Producer;

In particular, if

$$C(q) = r \cdot \frac{1}{1+k} \cdot q^{1+k}, r > 0, k > 0,$$

then we have Substitutes if $k > 1$ and Complements for $0 < k < 1$.

We shall see that these results follow from the fact (which we shall establish) that our "one-point forecaster is a *Blackwell IG.* An IG has a collection of available *information structures.* A structure (to be defined more formally below) specifies a signal set and the associated posteriors. For a Blackwell IG, the available structures can be partially ordered in such a way that moving from a given structure to a structure that precedes it in the ordering can never hurt any Producer, no matter what the prior probabilities and the payoff function may be. Moving to the second structure cannot lower the average (over all signals) of the Producer's highest attainable expected payoff. In our applications, moreover, it will be reasonable to say that the second structure requires more effort. So for a Blackwell IG, more effort never hurts the Producer. Suppose, for example, that the IG partitions the possible states θ and tells the Producer the set in which the current θ lies. Suppose the available partitionings can be ordered so that each is a refinement of its predecessor. A refinement can never damage the Producer, whatever the payoff function and the prior may be. So this IG is a Blackwell IG and the preceding Complements/Substitutes results hold for this IG, just as they do for the "one-point forecaster" we have just considered.

Similarly, an IG who samples is a Blackwell IG if he reports the entire sample to the Producer, since a larger sample can never hurt the Producer. Moreover, it is reasonable to say that a larger sample means more effort. We can modify the preceding example so that θ is not a price but is the unknown mean of a population of prices π, and the distribution of prices is known to belong to a one-parameter family, where the mean θ is the parameter. The Producer has a prior on the possible values of θ. The IG samples from the population of prices and reports the entire sample, say S, to the Producer, who chooses a quantity that maximizes the expected value of $\pi q - C(q)$, where the expectation is calculated for the distribution whose mean is $E(\theta|S)$. We then again get the preceding Complements/Substitutes results.

Let us now illustrate, on the other hand, a *Non-Blackwell* IG. Suppose that all prices θ in $[0, 1]$ are possible, and that θ is uniformly distributed on $[0, 1]$. This IG partitions $[0, 1]$ into n *equally probable* subintervals. All positive integers m are available to the IG and it is *not* true that every partitioning in the IG's collection is a refinement of some other partitioning in the collection. More IG effort means higher n. The IG finds the subinterval in which the true θ lies and his signal tells the Producer what that subinterval is. This IG is not a Blackwell IG because one can find a payoff function whose highest attainable expected value (averaging over the possible signals) drops for certain increases in n[8]. Nevertheless it turns out — using very different techniques than those used to get the preceding results — that if we again consider the cost function $C(q) = r \cdot \frac{1}{1+k} \cdot q^{1+k}$, then we again have Substitutes if $k > 1$ and Complements for $0 < k < 1$. It is again the case, moreover, that more IG effort benefits the Producer.

Next suppose the Producer is a monopolist with a linear demand curve that randomly rotates with its quantity intercept fixed. Price is $1-\theta q$ and given the IG's signal y, the Producer chooses a nonnegative quantity q which maximizes $E((1-\theta q) \cdot q - C(q) \mid y)$ on the set of all nonnegative quantities. Again assume that C is thrice differentiable and that $C(0) = 0, C' > 0, C'' > 0$. It now turns out that for our forecaster, and indeed for any Blackwell IG:

we have Complements if $C'''(q) < 0$ for all $q > 0$.

There is no general result for the case $C''' > 0$.

For a final preview, let the Producer be a price-taker who is uncertain about cost. Given the IG's signal y, the Producer chooses a nonnegative quantity q which maximizes $E(q - \theta C(q) \mid y)$ on the set of all nonnegative quantities. Assume $C(0) = 0, C' > 0, C'' > 0$. Then, for our forecaster, and for any other Blackwell IG as well:

we have Complements if $\frac{1}{C'(q)}$ is convex on $[0, \infty)$.

We would have Substitutes if $\frac{1}{C'(q)}$ were concave on $[0, \infty)$, but there is no function C with that property.

Results such as those we have sketched may have implications for future empirical work on the complex relation between information technology advances, workforce shifts, and productivity. Our exceedingly brief glance at the present status of this work (in Section 1.1 above) suggests that the work is in a fluid state with room for new ideas. One new idea might be to study IG/Producer pairs, as in our results. Perhaps it is possible, for example, to identify and count information workers who provide forecasts, or information workers who do sampling, or even information workers who partition the set of states into equal-probability intervals. Perhaps it is possible to identify and count as well the workers who belong to the organizations that use the information collected by the information workers of a given type in making the organization's production decisions. If the theory concludes (under appropriate simplifying assumptions) that more effort by the IGs (information workers) of a given type leads the Producers who use the services of those IGs to increase their production quantities (Complements), then the theory

[8]See details in sectionB.

has presented empirical researchers with a hypothesis. Perhaps the hypothesis could then be tested, using the observations on information workers and information users which the empirical researchers have assembled.

1.3 Related literature.

What is novel about the main question we ask is that it does not concern the *maximum* of the expected payoffs a quantity-chooser who responds to signals can obtain. Rather we study the *maximizer* — the best quantity itself and how it changes when the set of signals changes. There are many papers that concern the value of an information source to a decision-maker — the decision-maker's highest attainable expected payoff when he responds to the source's signals. Such papers study the relation between value and various interesting properties of the source or of the decision-maker's payoff function. Typically they do not focus on the relation between such properties and the decision-maker's value-maximizing choices themselves.

Nevertheless, for the models we study, and the definitions of effort that we shall consider, we usually impose the reasonable requirement that the Producer benefits (or does not suffer) when the IG increases his efforts. Without imposing that requirement, we can still study the change in Producer's average best quantity when the signal set of a low-effort IG is replaced by the signal set of a high-effort IG. But it is reassuring to confirm that the requirement is indeed met. So the value-of-information literature is relevant.

For our purposes, a key piece of that literature consists of the value-of experiments papers by Blackwell (1951, 1953) and his various followers in mathematical statistics, as well as the economic-theory papers that build on the Blackwell results. (In Section 1.6 below we cite some of the main papers in each group). As already noted, a Blackwell IG's structures can be ranked so that a structure is at least as useful to every Producer as a structure with lower rank, no matter what the prior and the payoff function may be. The Blackwell theorems tell us that this statement is equivalent to another statement: the expected value (over the possible signals) of any convex function on the posteriors is not less for the higher-ranked structure. If a higher-ranked structure requires more effort, according to an appropriate definition of effort, then (for that definition) we have Complements (Substitutes) whenever the Producer's best quantity is a convex (concave) function of the posterior.

One might seek a connection between our problem and an early literature on the choices made by a monopolist who is uncertain about demand and is not risk-neutral Sandmo (1971), Leland (1972), Holthausen (1976)). These papers study the relation between the monopolist's risk preferences and his quantity and input choices. We study, by contrast, the quantity responses of a risk-neutral Producer to changes in the information on which he bases his production choices. The two agendas are distinct. It appears that the results of the earlier literature cannot be used in finding answers to our main question.

Similarly one might seek a connection between our problem and an oft-cited paper by Athey and Levin (Athey and Levin, 2002). The paper studies expected-payoff-maximizing decision-makers whose payoff depends on an action in $\mathbb{R}$ and a state of the world in $\mathbb{R}$. The decision-makers receive signals in $\mathbb{R}$ from one of two sources. A source is defined by a joint state/signal distribution. Some decision-makers prefer the first source and some prefer the second. The payoff functions are assumed to obey a key condition: there is a monotone best-action function on the set of possible signals. If the condition is met, then the family of decision-makers who prefer a given source can be characterized in a compact way.

In our setting, one can imagine a Producer with a fixed prior who ranks two alternative services proposed by an IG. They differ with respect to an appropriate measure of effort. Since the prior is fixed, each service is defined by a signal distribution for every state. Given a price of effort, the Producer would choose the service which best balances IG cost and expected payoff. A Complements/Substitutes question could then be posed: if the price of effort drops, will the Producer's average best quantity, when he uses his preferred service, rise or fall? The Athey/Levin paper is concerned, however, with the value of each service to a user. It is not concerned with the change in the average of the user's best quantity over the service's signals when we move from one service to the others.

1.4 The formal framework.

We start with the concept of an *experiment on the state set* Θ. For every state θ in Θ, an experiment generates a signal y in a set Y of possible signals. The signal is a random variable. For every θ in Θ, the experiment specifies the signals' probability distribution; a signal probability is often called a *likelihood*. More precisely, for each $\theta \in \Theta$, we have a *likelihood measure space*

$$L_\theta = (Y, \mathcal{Y}, \lambda_\theta),$$

where $\mathcal{Y}$ is a σ-algebra over Y, and λ_θ is a probability measure on $\mathcal{Y}$. (If Y is finite then $\mathcal{Y}$ is the set of all subsets of Y). An experiment on Θ is a collection

$$\mathcal{E} = \{L_\theta\}_{\theta\in\Theta}.$$

Note that *no* prior beliefs about the states are specified when we define an experiment.

Now suppose that we *are* given prior beliefs about the states. The beliefs, together with the experiment, imply:

- a posterior measure on Θ for every signal $y \in Y$
- a marginal distribution on the signal set Y.

The collection of posterior measures (one for each signal), together with the marginal measure on the signal set, will be called *the information structure associated with the experiment and the* *prior beliefs*. We will suppose that the Producer has prior beliefs. A signal, conveyed by the IG to the Producer, then identifies a posterior, and the Producer uses that posterior to choose an expected-payoff-maximizing production quantity. Whether or not the IG has prior beliefs is not pertinent to our problem. But for a given level of IG effort, it will be convenient to talk of the "IG's set of possible posteriors" as well as "the IG's set of possible signals". One interpretation of the former phrase is that the IG indeed shares the Producer's prior beliefs and can therefore compute the set of possible posteriors.

Given the Producer's payoff function, an information structure has a *value* for the Producer, namely the average, over all the possible posteriors (signals), of the highest attainable expected payoff under the posterior. The experiment is our model of the IG's activities. The Producer is not directly concerned with those activities. What concerns him is the value of the structure associated with the experiment. What concerns us is the value of the structure as well as the average, over all the structure's signals, of the Producer's expected-payoff-maximizing quantity.

To state formally the relation between the prior beliefs, the experiment $\mathcal{E} = \{L_\theta\}_{\theta\in\Theta}$ (where $L_\theta = (Y, \mathcal{Y}, \lambda_\theta))$, and the associated information structure, let the prior beliefs be expressed by a measure space $(\Theta, \mathcal{T}, G)$ where $\mathcal{T}$ is a σ-algebra on Θ and G is a probability measure on $\mathcal{T}$. Then G implies a (marginal) signal-probability measure W_G on $\mathcal{Y}$ as well as a posterior state-probability measure F_{yG} on $\mathcal{T}$ for every signal $y \in Y$. The pair $(\Theta, \mathcal{T})$ will be fixed throughout the discussion. The pair $(Y, \mathcal{Y})$ may change from one experiment to another, but it is cumbersome to keep repeating the symbol $\mathcal{Y}$. So we will define the information structure associated with G and the experiment $\mathcal{E} = \{(Y, \mathcal{Y}, \lambda_\theta)\}_{\theta\in\Theta}$ as a triple

$$I = (Y, \{F_{yG}\}_{y\in Y}, W_G).$$

It is understood that F_{yG} is a measure on $\mathcal{T}$ and that W_G is a measure on the *sigma*-algebra $\mathcal{Y}$ which appears in the definition of the experiment $\mathcal{E}$. When Y is finite, $\mathcal{Y}$ is the set of all subsets of Y. It will often be the case that G is understood. Then we drop the subscript G and we write the information structure associated with the experiment $\mathcal{E}$ as

$$I = (Y, \{F_y\}_{y\in Y}, W).$$

That triple contains all the Producer needs to know in order to compute the value of the structure (given his payoff function), his best quantity for every signal, and the average of his best quantities over all the signals.

1.5 Four Blackwell theorems

In several of the main classes of experiments that we will be studying, our results will rest in part on one of four *Blackwell theorems*. In the first of them the state set and the signal set are finite. In the second theorem only the state set is required to be finite. In the third theorem neither set is required to be finite. The fourth theorem obtains some but not all of the third theorem's results. It uses part of the third theorem's proof but requires fewer assumptions. It will be especially useful for the Complements/Substitutes question.

In each theorem we start by fixing the state set Θ. Each theorem concerns two experiments on Θ, denoted $\mathcal{E}$ and $\mathcal{E}'$. Informally, one may think of $\mathcal{E}'$ as the "better" of the two. Each theorem asserts the equivalence of three statements, which we first describe informally:

(a) *$\mathcal{E}'$ is at least as informative as $\mathcal{E}$. (Any expected payoff that can be attained for $\mathcal{E}$ can also be attained for $\mathcal{E}'$).* First fix a set of actions and a payoff function on the action/state pairs. Then, for each experiment, consider all the possible action-choosing rules, where each rule assigns an action to each of the experiment's signals. For every state, a rule determines an expected payoff when we average over the signals. So for every rule, we have a function from states to expected payoffs, and for the set of all possible rules we have a set of such state-contingent expected-payoff functions. Finally, consider the set of all probability mixtures of those functions. We say that $\mathcal{E}'$ is at least as informative as $\mathcal{E}$ if the set of mixtures for $\mathcal{E}'$ includes the set of mixtures for $\mathcal{E}$. That implies the following weaker statement. For any prior on the states and any payoff function, consider the expected value, over all the signals, of the highest attainable expected payoff given the signal; that expected value is not lower for $\mathcal{E}'$ than for $\mathcal{E}$. The weaker statement is used in the first of the three theorems.

(b) $\mathcal{E} = (Y, \mathcal{Y}, \lambda_\theta)$ *is a garbling of* $\mathcal{E}' = (Y', \mathcal{Y}', \lambda'_\theta)$. For a fixed state, say θ^*, and any fixed subset H of the experiment $\mathcal{E}$'s signal set Y, consider the experiment-$\mathcal{E}$ probability $\lambda_{\theta^*}(H)$. That probability equals the average, under a suitable randomization, of the values taken by experiment $\mathcal{E}'$'s measure λ'_{θ^*} on the subsets of that experiment's signal set Y'.

(c) *A statement about convex functions of the posteriors.* For any convex sfunction ϕ on the set of possible distributions on Θ, and any prior on Θ, consider the expected value of ϕ, over all the posteriors that are implied by the experiment's signals. That expected value is not less for $\mathcal{E}'$ than for $\mathcal{E}$.

It is the finite first theorem that is most familiar in the economic literature. The most complete treatment of the finite case remains Marschak and Miyasawa (1968). The best-known portion of the first theorem asserts the equivalence of statements (a) and (b), though the equivalence of all three statements is shown in Marschak and Miyasawa. To prove that (a) implies (b), one has to show that the required garbling probabilities exist. Several papers by economists provide alternative ways of doing so.[9]

In Blackwell's two fundamental papers, the signal set is not required to be finite, but the state set is. Once the finiteness of the state set is dropped, serious measure-theoretic difficulties arise in constructing an appropriate version of the three statements and a proof of their equivalence.

[9]McGuire (1972), Ponssard (1975), Crémer (1982).

That is disturbing, since in economic settings it is often natural to model the state set as a continuum.

In mathematical statistics a substantial comparison-of-experiments literature followed Blackwell's two papers. Various approaches to dropping the finiteness of the state set were pursued. It is difficult, however, to extract appropriate results from this literature and to assemble them into a unified theorem which allows an arbitrary state set and shows the equivalence of (a), (b), and (c).[10] Fortunately we now have a unified theorem, prepared for an economic-theory audience. It is found in a doctoral dissertation by Zhixiang Zhang (2008). Zhang first presents the theorem in which (as in the Blackwell papers) the signal set is arbitrary but the state set is finite. He then drops the finiteness of the state set but partitions the state set into a finite number of sets. A sequence of finite partitionings, each a refinement of its predecessor, is considered and the finite-state-set theorem is applied to each member of the sequence. Using a limiting procedure, the desired theorem is obtained.[11]

The first of the four Blackwell theorems that we now present is the one familiar in economics, where both signal and state sets are finite. The second is Zhang's starting place — the signal set is arbitrary, as in the Blackwell papers, but the state set is finite. The third is the fully general theorem, where both sets are arbitrary. (Zhang's theorem shows the equivalence of five statements, but we omit two of them). The fourth theorem obtains some of the third theorem's results but with fewer assumptions.

Before stating the theorems, we develop formal versions of statements (a), (b), and (c) and the terms that are used in those statements. They are presented in a general way, but they take a simpler form in the first Theorem, where state set and signal set are both finite. We start with "at least as informative as".

Let $A \subset I\!R$ be an *action set*. Consider an action-taker — our Producer, for example — who responds to every signal in the signal set Y of the experiment $\mathcal{E} = (Y, \mathcal{Y}, \{\lambda_\theta\}_{\theta\in\Theta})$ by choosing an action in A. To describe the set of action-choosing rules in the most general way, consider a measurable *action space* $(A, \mathcal{A})$. Then define the set of possible action-choosing rules as

$$D(\mathcal{E}, A) = \{\delta \ : \ \delta \text{ is a measurable function from } Y \text{ to } A\}.$$

Now let

$$u : A \times \Theta \to I\!R$$

[10]A brief survey of the post-Blackwell literature is LeCam (1996). Among the papers mentioned is Strassen (1965), which proves the equivalence of (b) and (c) under certain assumptions. Some of the work surveyed by LeCam is found in a lengthy monograph by Torgerson (1991). An earlier version of some of that monograph's results are found in Torgerson (1976). Other key papers are LeCam (1964), Lehmann (1955), and Lehmann (1988).

[11]Amershi (1988) also develops a non-finite Blackwell theorem for a primarily economic audience. In that theorem one finds the equivalence of statements (a) and (b), but the theorem is silent on statement (c). The techniques are quite different than those used by Zhang.

be a payoff function. For the state θ, the rule $\delta \in D(\mathcal{E}, A)$, the payoff function u, and the experiment $\mathcal{E} = (Y, \mathcal{Y}, \{\lambda_\theta\}_{\theta\in\Theta})$, consider

$$U(\theta, u, \delta, \mathcal{E}) = \int_Y u(\delta(y), \theta)\lambda_\theta(dy).$$

If we fix $u, \delta, \mathcal{E}$, then $U(\ \cdot \ , u, \delta, \mathcal{E})$ is a *state-contingent expected-payoff function*: it assigns to every state θ the expected payoff when the action-choosing rule δ is used and expectation is taken over the possible signals of the experiment $\mathcal{E}$. For $\mathcal{E}$, the set of all state-contingent expected-payoff functions, when we allow the action-choosing rule to vary, is

$$B(u, Y, \mathcal{E}, A) \equiv \{U(\ \cdot \ , u, \delta, \mathcal{E}) \ : \ \delta \in D(\mathcal{E}, A)\}.$$

For $\mathcal{E}'$, the set is

$$B(u, Y, \mathcal{E}', A) \equiv \{U(\ \cdot \ , u, \delta, \mathcal{E}') \ : \ \delta \in D(\mathcal{E}', A)\}.$$

Finally, let $\bar{B}(u, Y, \cdot, A)$ denote the convex closure of the set $B(u, Y, \cdot, A)$, where closure is defined by pointwise convergence. We are now ready to define *at least as informative as*.

Definition

The experiment $\mathcal{E}' = (Y', \mathcal{Y}', \{\lambda'_\theta\}_{\theta\in\Theta})$ is *at least as informative as the experiment* $\mathcal{E} = (Y, \mathcal{Y}, \{\lambda_\theta\}_{\theta\in\Theta})$ if for every action space $(A, \mathcal{A})$ and every payoff function $u : A \times \Theta \to I\!R$, we have

$$\bar{B}(u, Y, \mathcal{E}, A) \subseteq \bar{B}(u, Y', \mathcal{E}', A).$$

Consider any state-contingent expected-payoff function obtainable in $\mathcal{E}$. If $\mathcal{E}'$ is at least as informative as $\mathcal{E}$, then by using a suitable probability mixture of the action-choosing rules in $D(\mathcal{E}', A)$, we can exactly duplicate that function in the finite case and we can approximate it as closely as desired in the non-finite case.

Next we define garbling. The definition uses Markov kernels. A *Markov kernel from* $(Y', \mathcal{Y}')$ *to* $(Y, \mathcal{Y})$ is a collection of measures $\{p_{y'}\}_{y'\in Y'}$ such that for each $y' \in Y'$ the following holds: (i) $p_{y'}$ is a probability measure on $(Y, \mathcal{Y})$; and (ii) for any set $H \in \mathcal{Y}$ the function $p_{y'}(H) : Y' \to [0, 1]$ is measurable with respect to $(Y', \mathcal{Y}')$.

Definition.

The experiment $\mathcal{E} = (Y, \mathcal{Y}, \{\lambda_\theta\}_{\theta\in\Theta})$ is a *garbling* of the experiment $\mathcal{E}' = (Y', \mathcal{Y}', \{\lambda'_\theta\}_{\theta\in\Theta})$ if the following hold:

(i) There exists a Markov kernel $\{p_{y'}\}_{y'\in Y'}$ from $(Y', \mathcal{Y}')$ to $(Y, \mathcal{Y})$.

(ii) For every $\theta \in \Theta$ and every set $H \in \mathcal{Y}$ we have $\lambda_\theta(H) = \int_{y'\in Y'} p_{y'}(H)\lambda'_\theta(dy')$.

We shall say that *the information structure associated with the experiment $\mathcal{E}$ and the prior G is a garbling of the information structure associated with the experiment $\mathcal{E}'$ and G* if $\mathcal{E}$ is a garbling of $\mathcal{E}'$.

Finally, to state formally our condition about a convex or concave function on the posteriors, first consider the prior beliefs about the states. (These might be held, for example, by our Producer). Using the notation already introduced in Section 1.4, the beliefs are expressed by the measure space

$$(\Theta, \mathcal{T}, G),$$

where G is a probability measure on $\mathcal{T}$. For every signal y in the experiment $\mathcal{E}$'s signal set Y, there is a posterior measure F_{yG} on $\mathcal{T}$. (Recall that when G is understood, we omit the second subscript and use the symbol F_y). Let $\Delta(\Theta)$ denote the set of all probability measures on $\mathcal{T}$. (If Θ is finite, with ℓ states, then $\Delta(\Theta)$ is the standard $\ell - 1$-simplex in $\mathbb{R}^\ell$).

Consider a function ϕ from $\Delta(\Theta)$ to the reals and consider the random variable $\phi(F_{yG})$. Its distribution depends on the marginal distribution of the signals $y \in Y$. We shall let $E_{y\in Y}\ \phi(F_{yG})$ denote the expected value of ϕ. Thus

$$E_{y\in Y}\ \phi(F_{yG}) = \int_Y\ \phi(F_{yG}) W_G(dy).$$

The third of our three conditions concerns the expected value of ϕ when ϕ is convex. It compares the expected value of ϕ for the experiment $\mathcal{E}$ with its counterpart for the experiment $\mathcal{E}'$, namely $E_{y'\in Y'}\ \phi(F'_{y'G})$, where $F'_{y'G}$ denotes the posterior measure on $\mathcal{T}$ implied by the experiment $\mathcal{E}$'s signal $y' \in Y$.

In the first of the three theorems, where state set and signal sets are finite, "at least as informative as" (in statement (a) of the theorem) takes a weaker form than the set-inclusion relation defined above. Moreover "garbling" is given a matrix definition which is, as we shall see, a special case of the Markov-kernel garbling defined above.

First Blackwell Theorem

Suppose that $\Theta = \{\theta_1, \ldots, \theta_i, \ldots, \theta_n\}$. Let $g_1, \ldots, g_n$, with $g_i > 0, i = 1, \ldots, n$, be prior probabilities on Θ. Consider the experiment $\mathcal{E} = (Y, \mathcal{Y}, \{\lambda_\theta\}_{\theta\in\Theta})$, where $Y = \{y_1, \ldots, y_j, \ldots, y_m\}$, $\mathcal{Y}$ is the set of all subsets of Y, and λ_θ denotes the likelihood vector $\left(\lambda_\theta^1, \ldots, \lambda_\theta^j, \ldots, \lambda_\theta^m\right)$. Consider also the experiment $\mathcal{E}' = (Y', \mathcal{Y}', \{\lambda'_\theta\}_{\theta\in\Theta})$, where $Y' = \{y'_1, \ldots, y'_j, \ldots, y'_{m'}\}$, $\mathcal{Y}'$ is the set of all subsets of Y', and λ'_θ denotes the likelihood vector $\left(\lambda'^1_\theta, \ldots, \lambda'^j_\theta, \ldots, \lambda'^{m'}_\theta\right)$.

The following three statements are equivalent:

(a) Let $\pi^y = (\pi_1^y, \ldots \pi_n^y)$, and $\pi'^{y'} = (\pi_1'^{y'}, \ldots \pi_n'^{y'})$ be, respectively, the vector of posterior state probabilities implied by experiment $\mathcal{E}$'s signal y and the prior probabilities and by experiment $\mathcal{E}'$'s signal y' and the prior probabilities. Consider a payoff function $u : A \times \Theta \to I\!R$. Let $u^*(y)$ be the highest attainable expected value of u under the posterior-probability vector π^y, and let $u^{*'}(y')$ be the highest attainable expected value of u under the posterior-probability vector $\pi'^{y'}$. The expected value of $u^{*'}(y')$ over the signals y' in Y' is not less than the expected value of $u^*(y)$ over the signals y in Y.

(b) Let Λ_{ij} denote $\lambda_{\theta_i}^j$ and let Λ'_{ij} denote $\lambda'^j_{\theta_i}$. Consider the two likelihood matrices $\Lambda = ((\Lambda_{ij}))$ and $\Lambda' = \left((\Lambda'_{ij})\right)$. There exists a row-stochastic m'-by-m "garbling" matrix B such that

$$\underbrace{\Lambda}_{n\text{-by-}m} = \underbrace{\Lambda'}_{n\text{-by-}m'} \cdot \underbrace{B}_{m'\text{-by-}m} .$$

(c) For every real-valued convex function ϕ on the n-simplex $\Delta(\Theta)$, we have:

$$E_{y \in Y'} \ \phi\left(\pi'^{y'}\right) \geq E_{y \in Y} \ \phi\left(\pi^y\right).$$

The matrix definition of garbling in Statement (b) is a special case of our general Markov-kernel definition of garbling. To see this, observe that Λ_{ij} is experiment-$\mathcal{E}$'s probability of the singleton $\{y_j\}$, a subset of Y; and $\lambda'_{i\ell}$ is experiment-$\mathcal{E}'$ probability of the singleton $\{y'_\ell\}$, a subset of Y'. Now let $\mathcal{Y}$ be the set of all subsets of Y and let $\mathcal{Y}'$ be the set of all subsets of Y'. For each $y'_j \in \{y'_1, \ldots, y'_{m'}\}$ we have a probability measure on $(Y, \mathcal{Y})$, namely $p_{y'_j}$, which belongs to the Markov kernel $\{p_{y'}\}_{y' \in Y'}$. (The Markov-kernel requirement that for fixed $H \in \mathcal{Y}$, the function $p_{y'}(H) : Y' \to [0,1]$ be measurable with respect to $(Y', \mathcal{Y}')$ is met automatically, since $\mathcal{Y}'$ is the set of all subsets of Y'). For the singleton $\{y_r\} \in \mathcal{Y}$, we have (using the above "garbling matrix" B)

$$p_{y'_j}(\{y_r\}) = B_{jr},$$

and for any set $H \subseteq Y$, we have

$$p_{y'_j}(H) = \sum_{y_t \in H} p_{y'_j}(\{y_t\}).$$

For the state $\theta = \theta_k$ and the singleton $\{y_r\} \in \mathcal{Y}$, the equality in part (ii) of our general (Markov-kernel) definition of garbling becomes:

$$\lambda_{\theta_k}(\{y_r\}) = \sum_{j=1}^{m'} \left[p_{y'_j}(\{y_r\}) \cdot \lambda'_{\theta_k}(\{y'_j\}) \right] = \sum_{j=1}^{m'} B_{jr} \cdot \lambda'_{kj}.$$

That is precisely what the equality $\Lambda = \Lambda' \cdot B$ tells us. For any set $H \subseteq Y$ we have

$$\lambda_{\theta_k}(H) = \sum_{y_t \in H} \lambda_{\theta_k}(\{y_t\}).$$

In our next theorem, the state set remains finite but the signal sets need not be.

Second Blackwell Theorem

Suppose the state set Θ is finite. Consider two experiments, $\mathcal{E} = (Y, \mathcal{Y}, \{\lambda_\theta\}_{\theta\in\Theta})$ and $\mathcal{E}' = (Y', \mathcal{Y}', \{\lambda_\theta\}_{\theta\in\Theta})$ and a prior probability measure space $(\Theta, \mathcal{T}, G)$ on the states such that every state has positive mass. The following three statements are equivalent:

(a) $\mathcal{E}'$ is at least as informative as $\mathcal{E}$.

(b) $\mathcal{E}$ is a garbling of $\mathcal{E}'$.

(c) For every continuous real-valued convex function ϕ on $\Delta(\Theta)$, we have

$$E_{y'\in Y'}\ \phi(F'_{y'G}) \geq E_{y\in Y}\ \phi(F_{yG}),$$

where F_{yG} is the posterior given the experiment $\mathcal{E}$-signal $y \in Y$ and $F'_{y'G}$ is the posterior given the experiment $\mathcal{E}'$-signal $y' \in Y'$.

In the third theorem, the state set Θ need no longer be finite. We add two assumptions on the experiments $\mathcal{E}', \mathcal{E}$ which the theorem compares.

Assumption A1

Consider the experiment $\mathcal{E}' = (Y', \mathcal{Y}', \{\lambda'_\theta\}_{\theta\in\Theta})$. There exists a σ-finite measure ζ such that for every $\theta \in \Theta$, the measure λ'_θ is absolutely continuous with respect to ζ.

Assumption A2

Consider the experiment $\mathcal{E}' = (Y', \mathcal{Y}', \{\lambda'_\theta\}_{\theta\in\Theta})$, where Θ is a complete separable metric space. Consider also the experiment $\mathcal{E} = (Y, \mathcal{Y}, \{\lambda_\theta\}_{\theta\in\Theta})$, where $\mathcal{Y}$ is the Borel σ-algebra on Y.

- If q is any bounded measurable function on Y', then $\int_{Y'}\ q(x)\lambda'_\theta(dx)$ is continuous in θ.
- If r is any bounded continuous function on Y, then $\int_{Y}\ q(x)\lambda_\theta(dx)$ is continuous in θ.

Note that A1 is automatically satisfied when Θ is finite(let ζ be $\sum_{\theta\in\Theta}\ \lambda'_\theta$). Condition A2 is also automatically satisfied when Θ is finite.

Third Blackwell Theorem (Zhang)

Let Θ be a complete separable metric space. Consider the experiment $\mathcal{E}' = (Y', \mathcal{Y}', \{\lambda_\theta\}_{\theta\in\Theta})$ and the experiment $\mathcal{E} = (Y, \mathcal{Y}, \{\lambda_\theta\}_{\theta\in\Theta})$, where Y, Y' are compact metric spaces, $\mathcal{Y}$ is the Borel σ-algebra on Y, and $\mathcal{Y}'$ is the Borel σ-algebra on Y'. Consider a prior probability measure space $(\Theta, \mathcal{T}, G)$ on the states such that $G(S) > 0$ for every nonempty open subset S of Θ. Suppose that $\mathcal{E}'$ satisfies Assumption A1 and that $\mathcal{E}, \mathcal{E}'$ satisfy Assumption A2. Then the following three statements are equivalent:

(a) $\mathcal{E}'$ is at least as informative as $\mathcal{E}$.

(b) $\mathcal{E}$ is a garbling of $\mathcal{E}'$.

(c) For every continuous real-valued convex function ϕ on $\Delta(\Theta)$, we have

$$E_{y' \in Y'} \ \phi(F'_{y'G}) \geq E_{y \in Y} \ \phi(F_{yG}),$$

where F_{yG} is the posterior given the experiment $\mathcal{E}$ signal $y \in Y$ and $F'_{y'G}$ is the posterior given the experiment $\mathcal{E}'$ signal $y' \in Y'$.

We now turn to the final Fourth Blackwell Theorem. This theorem does not require Assumption A2. That assumption is violated in an important case: the state set is a continuum (e.g., the unit interval), the signal set is finite, and each signal identifies a set in a fixed partitioning of the state set. Fortunately, in Zhang's proof of the Third Theorem, Assumption A2 is only used in arguing that statement (c) implies statement (b). In our Complements/Substitutes problems we are primarily interested in the equivalence of (a) and (b), and the fact that (b) implies (c). Those require Assumption A1 but not Assumption A2. Accordingly we use part of Zhang's proof to establish the following theorem.

Fourth Blackwell Theorem

Let Θ be a complete separable metric space and consider the experiments $\mathcal{E}', \mathcal{E}$, and the prior probability measure space, discussed in the Third Theorem. Suppose that $\mathcal{E}'$ satisfies Assumption A1. Then statements (a) and (b) are equivalentss and statement (b) implies statement (c).

1.6 Plan of the rest of the paper.

Section 2 concerns Blackwell IGs, whose structures can be ranked so that for every Producer, a higher-ranked structure is at at least as valuable as a lower-ranked one. Section 2.1 summarizes the way we get Complements/Substitutes results for Blackwell IGs using the Blackwell theorems, and also summarizes an alternative path to these results for the special class of Blackwell IG studied in 2.3. Section 2.2 concerns what we shall call "strict garbling" of finite structures. We state two strict-garbling theorems, which permit us to obtain *strict* Complements/Substitutes results for certain Blackwell IGs and certain Producers.

The Blackwell IG in Section 2.3 has a family of finite information structures, indexed by $k \in [0, 1]$. The typical structure is a linear combination of two anchor structures. One of them is more useful to every Producer than the other. The superior anchor structure is given weight k and higher k means more IG effort. It turns out that the "one-point sforecaster", discussed above in the first of our previews, is a special case.

In Section 2.4 we obtain counterparts of the finite-structure results in 2.3 for the case where the state set and the signal sets are continua.

Section 2.5 studies another Blackwell IG, whose structures are again indexed by $k \in [0, 1]$. This time k is the probability that the IG's structure is the "better" of two anchor structures.

Section 2.6 turns to a general question: when is a Producer's best quantity a convex (concave) function of the posterior? Sufficient conditions for convexity (concavity) are obtained.

Section 3 moves to a non-Blackwell IG. The Blackwell Theorems are not used in this section. In section 3.1 the states θ have support $[0, 1]$ Two priors are considered. One of then is gven by $G(\theta) = \theta^{\frac{1}{a}}, a > 0$ (a generalization of the uniform distribution). The other is a generalization of the Pareto-Levy distribution. The IG partitions $[0, 1]$ into n equal-probability sets and higher n means more effort. A Producer whose best quantity depends on the mean of the posterior is considered. For a class of best-quantity functions it is shown — for both priors — that more effort benefits the Producer. Moreover there is a subclass for which we have Complements and another subclass for which we have Substitutes. The same results are obtained again if we suppose that the prior is again $G(\theta) = \theta^{\frac{1}{a}}$, with $a > 0$ and the IG partitions $[0, 1]$ into subintervals of equal *length.* The proofs turn out to be intricate and specialized. That may lead one to suspect that the results might be obtained instead as special cases of some more general theorem with very weak assumptions about the prior. Section 3.2 deals with this question. We first show that a suggestive general theorem from the theory of stochastic orders does not yield our results. We then consider what our results tell us about two equal-probability intervals versus three. For a Producer with a certain payoff function three is better than two and we have Complements (average best quantity rises when we go from two to three). We construct a prior distribution on $[0, 1]$ for which three continues to be better than two but we now have Substitutes. That supports the conjecture that once we leave the Blackwell IG, Complements/Substitutes results require special assumptions and special techniques. When we vary the prior and the payoff function, each case presents its own peculiar challenges.

Section 3.3 summarizes results obtained in another paper for a non-Blackwell IG and a Producer who is an inventory manager. The unknown θ is tomorrow's demand. Payoff equals $\min(q, \theta) - cq$, where q is today's order and c, with $0 < c < 1$, is a fixed unit cost. The Producer has a prior on θ and responds to the IG's signal with an expected-payoff-maximizing order. The IG has a "base" distribution on the states. Each of his signals determines a scale/location transform of the base distribution and that transform becomes the Producer's posterior. More IG effort means a smaller average scale parameter. Complements/Substitutes results are obtained, both for the *ex ante* quantity q and the *ex post* quantity $\min(q, \theta)$, where q is the optimal order. Average *ex post* quantity rises when the IG works harder. But average *ex ante* quantity rises whenever the IG works harder if c is above a critical value and falls when c is below that value. Again, the proof techniques are peculiar to this particular IG/Producer pair.

Section4 presents Complements/Substitutes applications of our theorems. They include the results we have previewed. Section 5 concludes, briefly sketching some of the many paths that lie open for future research.

All the technical proofs are relegated to the appendix.

1.7 Notation and terminology.

We use two alternative symbols for expectation. If T is a distribution (measure), then E_T denotes expectation under T. If a probability distribution (measure) with support X is understood, then $E_{x \in X}$ will be an alternative symbol for expectation.

Now let F be a distribution on Θ (or, more generally, a probability measure on Θ). Let $u : A \times \Theta \to I\!R$ be a payoff function for which the expected value $E_F\ u(q, \theta)$ has a largest maximizer in A. Then we will call that payoff function *regular for* F. We let $\hat{q}(F; u)$ denote the largest maximizer and we refer to $\hat{q}(\ \cdot\ ; u)$ as *the best-quantity function for* (F, u). We let $\tilde{V}_u(F)$ denote $E_F\ u(\hat{q}(F; u), \theta) = \max_q (E_F\ u(q, \theta))$. We call $\tilde{V}_u(F)$ *the value of* F *for the payoff function* u.

Suppose we are given a structure $I = (Y, \{F_y\}_{y \in Y}, W)$. We will call the payoff function U *regular for the structure* I if u is regular for every F_y with $y \in Y$. For a regular function u, the symbol $Q_u(I)$ will denote $E_{y \in Y}\ \hat{q}(F_y; u)$, and $V_u(I)$ will denote $E_{y \in Y}\ \tilde{V}_u(F_y)$. We will call $V_u(I)$ *the value of the structure* I *for the payoff function* u.

2 The Blackwell Information-gatherer

2.1 Obtaining Complements/Substitutes results with and without the Blackwell theorems.

We have already commented , in the "Preview" section 1.2 and the "Related literature" section 1.3, on the use of statement (c) of the Blackwell theorems, concerning convex functions on the posteriors. Statement (c) is an important tool in obtaining Complements/Substitutes results for a Blackwell IG. To use it we first have to show that the IG whom we are modeling is indeed a Blackwell IG. We have to show that we can partially order the IG's structures so that statement (a) holds when we compare a given structure with a lower-ranking one. Moreover we seek a reasonable definition of the IG's effort such that more effort coincides with "at least as informative as".

One way to show that statement (a) holds is to exhibit explicitly the Markov kernel required in Statement (b). That is seldom done in the economics literature, even when both signal and state sets are finite. We shall do so twice for a Blackwell IG who chooses a mixture of two "anchor" structures, first in Section 2.3.4 and then again in 2.4. In the first variant, signal and state sets are finite. In the second variant they are not. Exhibiting the Markov kernel explicitly allows us to interpret in a new way the superiority of a high-effort structure over a low-effort structure. We explicitly see the way that the low-effort signals distort the information conveyed by the high-effort signals. Once statement (b) is established for two structures, we also have (using statement (c)) a Complements (Substitutes) result for the two structures if the Producer's best-quantity function is convex (concave) in the posterior.

It turns out that for the IG whose structures are a mixture of two anchor structures, we can obtain the same results without using the Blackwell theorems at all, even though this IG is a Blackwell IG. Each structure is indexed by $k \in [0, 1]$. In the alternative path to our results, the key finding is that for any convex (concave) function on the posteriors, the expected value of that function, over all of the k-structure's signals, is convex (concave) in k. That has additional implications. If a Producer has to pay for IG effort, then the cost of k has to rise sufficiently rapidly if the Producer's best k is going to lie in the interior of $[0, 1]$. Moreover if best quantity is convex (concave) in the posterior, then the Complements (Substitutes) effect of a small increase in k is stronger for large k than for small k.

2.2 Strict garbling of finite structures

Statement **(c)** of the finite First Blackwell Theorem is particularly important for the Complements/Substitutes question. Suppose the state set is finite and the Producer uses an Information-gatherer whose structures have a finite signal set and form a Blackwell family. In the cases we consider it is reasonable to say that the first of two structures requires more effort than the second if the second is a garbling of the first. Then statement **(c)** tells us that we have Complements if the Producer's best quantity is convex in the posteriors and Substitutes if it is concave. But there are compelling examples in which both best quantity and highest attainable expected payoff are *strictly* convex, or *strictly* concave, in the posterior *mean*.[12] Since the mean is a linear function of the posterior itself, best quantity and highest attainable expected payoff are strictly convex (strictly concave) in the posterior itself.

The First Blackwell Theorem, as it stands, does not allow us to exploit such strictness. It only tells us that if two experiments $\mathcal{E}$ and $\mathcal{E}'$ have signal sets Y, Y' and likelihood matrices Λ, Λ' which obey the garbling condition $\Lambda = \Lambda' \cdot B$, where B is row-stochastic, then the average (over all signals) of a strictly convex best quantity cannot drop when the Information-gatherer goes from $\mathcal{E}$ to $\mathcal{E}'$. It would be very useful to modify the garbling condition in such a way that

[12] Suppose, for example, that the Producer's payoff is $u = \theta \cdot q - \frac{q^{\beta+1}}{1+\beta}$, with $\beta \in (1, 2)$. Then best quantity given the IG's signal, as well as highest attainable expected payoff given the signal, are strictly convex functions of the mean of the posterior distribution of θ which the signal defines.

"cannot drop" becomes "must rise" when we have strict convexity (and "cannot rise" becomes "must drop" when we have strict concavity).

We now provide such a modification. The garbling condition is replaced by a *strict* garbling condition. This requires that the row-stochastic garbling matrix B have a further property: it has at least one column with two positive entries, say the entry corresponding to the signal $y'_u \in Y'$ and the entry corresponding to the signal $y'_v \in Y'$. Without that further condition, B could be the identity matrix. Then $\Lambda = \Lambda'$, which would certainly exclude a "must rise" statement.

If we are interested in the case where best average payoff given a signal, or best quantity given a signal, are strictly convex (strictly concave) in the posterior itself (not its mean), then adding the further condition on B is not enough. We also have to place a full-rank condition on the likelihood matrix Λ'. We have to exclude matrices with identical columns. Otherwise we would be permitting the following situation:

$$\Lambda = \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1/3 & 2/3 \\ 1/3 & 2/3 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \Lambda' \cdot B.$$

Then in both experiments, the posterior defined by each signal equals the prior. Even for a "strictly convex" Producer, best quantity and average best payoff are the same for one experiment as for the other. Such situations are ruled out if we require that the likelihood matrix Λ' have full rank.

If, on the other hand, the best average payoff given a signal, or the best quantity given a signal, are strictly convex (strictly concave) functions of the *mean* of the posterior defined by the signal, then to obtain results about the average of those two variables over all signals, we require that the mean of the posterior defined by the signal y'_u be not equal to the mean of the posterior defined by the signal y'_v.

As before, we let the state set be $\Theta = \{\theta_1, \ldots, \theta_i, \ldots, \theta_n\}$ and we let the prior state probabilities be $p_1, \ldots, p_i, \ldots, p_n$, all of which we assume to be positive. We shall compare an experiment $\mathcal{E}$, in which the signal set is $Y = \{y_1, \ldots, y_s, \ldots y_m\}$ and the n-by-m likelihood matrix is $\Lambda = ((\lambda_{is}))$, with an experiment $\mathcal{E}'$, in which the signal set is $Y' = \{y'_1, \ldots, y'_t, \ldots y'_{m'}\}$ and the n-by-m' likelihood matrix is $\Lambda' = ((\lambda'_{it}))$. The (marginal) probability of signal $y_s \in Y$ is $\Pr(y_s) = \sum_{i=1}^n p_i \cdot \lambda_{is}$ and the (marginal) probability of signal $y_t \in Y'$ is $\Pr(y'_t) = \sum_{i=1}^n p_i \cdot \lambda'_{it}$. As before, all signal probabilities are assumed to be positive. We shall need the posterior state probabilities for each signal. Let q_{is}, q'_{it} denote, respectively, the posterior probability of state θ_i given signal $y_s \in Y$ and given signal $y'_t \in Y'$. We have:

$$q_{is} = \frac{p_i \cdot \lambda_{is}}{\sum_{i=1}^n p_i \cdot \lambda_{is}}; \; q'_{it} = \frac{p_i \cdot \lambda'_{it}}{\sum_{i=1}^n p_i \cdot \lambda'_{it}}.$$

We shall use the symbols

$$\vec{q_s} \equiv \begin{pmatrix} q_{1s} \\ \vdots \\ q_{ns} \end{pmatrix}, \vec{q'_t} \equiv \begin{pmatrix} q'_{1t} \\ \vdots \\ q'_{nt} \end{pmatrix}.$$

Now let $\Delta(\Theta)$ again denote the simplex $\{(p_1, \ldots, p_n) | p_i \geq 0 \ \forall i; \sum_{=1}^{n} \ p_i = 1\}$. Note that $\vec{q_s}, \vec{q'_t} \in \Delta(\Theta)$. For a given function $\phi : \Delta(\theta) \to I\!R$, define

$$V_\phi(\Lambda) = \sum_{s=1}^{m} \ (\Pr(y_s)) \cdot \phi(\vec{q_s}), \ V_\phi(\Lambda') = \sum_{t=1}^{m'} \ (\Pr(y'_t)) \cdot \phi(\vec{q'_t}).$$

We shall establish two theorems. The first theorem concerns a function ϕ of the posterior itself (not its mean), where ϕ is strictly convex or strictly concave; ϕ might be the Producer's best quantity given a signal, or it might be the Producer's best average payoff given a signal. The second theorem concerns a function of the mean of the posterior.

First strict-garbling theorem (a theorem about strict garbling and a strictly convex or strictly concave function on the posteriors.)

Suppose that we have

$$\Lambda = \Lambda' \cdot B,$$

where, given the prior state probabilities $p_1, \ldots, p_n$, the following conditions are satisfied by Λ, Λ', B:

(C1) $B = ((b_{ts}))$ is an m'-by-m row-stochastic matrix.

(C2) *(Strict garbling.)* At least one column of B has two positive entries, i.e., there exists j, u, v such that $u \neq v$ and $b_{uj} > 0, b_{vj} > 0$.

(C3) The rank of Λ' is m' (the columns of Λ' are linearly independent).

Then for any strictly convex (strictly concave) function $\phi : \Delta(\Theta) \to I\!R$ we have

$$V_\phi(\Lambda') > V_\phi(\Lambda) \quad (V_\phi(\Lambda') < V_\phi(\Lambda)).$$

.

Second strict-garbling theorem (a Theorem about strict garbling and a strictly convex or strictly concave function of the *mean* of the posterior.)

Suppose that we have

$$\Lambda = \Lambda' \cdot B,$$

where, given the prior state probabilities $p_1, \ldots, p_n$, the following conditions are satisfied by Λ, Λ', B:

(C1) $B = ((b_{ts}))$ is an m'-by-m row-stochastic matrix.

(C2) There exist j, u, v such that

$$(*) \qquad u \neq v \text{ and } b_{uj} > 0, b_{vj} > 0.$$

$$(**) \qquad E_{\vec{q'_u}}\ \theta \neq E_{\vec{q'_v}}\ \theta,$$

where $\vec{q'_u}, \vec{q'_v}$ are, respectively, the vector of posterior state probabilities associated with the signal y'_u and with the signal y'_v.

Then for any strictly convex (strictly concave) function $\phi : I\!R \rightarrow I\!R$, we have

$$V^*_\phi(\Lambda') > V^*_\phi(\Lambda) \quad (V^*_\phi(\Lambda') < V^*_\phi(\Lambda)),$$

where

$$V^*_\phi(\Lambda) \equiv \sum_{s=1}^{m} \ \Pr(y_s) \cdot \phi\left(E_{\vec{q_s}}\ \theta\right) \text{ and } V^*_\phi(\Lambda') \equiv \sum_{t=1}^{m'} \ \Pr(y'_t) \cdot \phi\left(E_{\vec{q'_t}}\ \theta\right).$$

In Section 4 below we illustrate the application of the second strict-garbling theorem.

2.3 Two Blackwell families of finite information structures, where each structure is defined by a parameter $k \in [0, 1]$.

2.3.1 A family in which each structure is a linear combination of two "anchor" structures: introduction.

A finite state set $\Theta = \{\theta_1, \ldots, \theta_n\}$ and a prior G on Θ with n positive prior state probabilities $g_1, \ldots, g_n$ are given. We study a family of experiments. In each experiment there are m signals, each with positive probability. Each experiment is indexed by a number k in $[0, 1]$. In the experiment $\mathcal{E}_k$, the signal set is $Y_k = \{y^k_1, \ldots, y^k_m\}$. Given a state, the likelihood of the jth signal in experiment $\mathcal{E}_k$, denoted Λ^k_{ij}, is a linear combination, with weight k, of the likelihood of the jth signal in each of two *anchor* experiments: the experiment $\mathcal{E}_1$ and the experiment $\mathcal{E}_0$. Thus for the state θ_i, the experiment-$\mathcal{E}_k$ likelihood of the signal y^k_j is

$$\Lambda^k_{ij} = k \cdot \Lambda^1_{ij} + (1 - k) \cdot \Lambda^0_{ij},$$

where Λ^1_{ij} is the experiment-$\mathcal{E}_1$ likelihood of the jth signal given the state θ_i, and Λ^0_{ij} is the experiment-$\mathcal{E}_0$ likelihood of the jth signal given the state θ_i. The likelihood matrix for experiment $\mathcal{E}_k$ will be denoted Λ_k. So

$$\Lambda_k = ((\Lambda^k_{ij})).$$

The following is the n-by-m matrix of *joint* signal/state probabilities for the experiment $\mathcal{E}_k$. In presenting the matrix, it is notationally convenient to let Λ'_{ij} denote Λ^1_{ij} and to let Λ_{ij} denote Λ^0_{ij}.

SIGNALS

$$\begin{array}{cc} & \begin{array}{cccc} \boldsymbol{y^k_1} & \boldsymbol{y^k_2} & & \boldsymbol{y^k_m} \\ \text{probability:} & \text{probability:} & \cdots & \text{probability:} \\ \sum_{i=1}^n g_i \cdot [k\Lambda'_{i1} + (1-k)\cdot\Lambda_{i1}] & \sum_{i=1}^n g_i \cdot [k\Lambda'_{i2} + (1-k)\cdot\Lambda_{i2}] & & \sum_{i=1}^n g_i \cdot [k\Lambda'_{im} + (1-k)\cdot\Lambda_{im}] \end{array} \\ \begin{array}{c} \text{S} \\ \text{T} \\ \text{A} \\ \text{T} \\ \text{E} \\ \text{S} \end{array} \begin{array}{c} \theta_1 \\ [\text{prob. } g_1] \\ \theta_2 \\ [\text{prob. } g_2] \\ \vdots \\ \theta_n \\ [\text{prob. } g_n] \end{array} & \left(\begin{array}{cccc} g_1 \cdot [k\Lambda'_{11} + (1-k)\cdot\Lambda_{11}] & g_1 \cdot [k\Lambda'_{12} + (1-k)\cdot\Lambda_{12}] & \cdots & g_1 \cdot [k\Lambda'_{1m} + (1-k)\cdot\Lambda_{1m}] \\ g_2 \cdot [k\Lambda'_{21} + (1-k)\cdot\Lambda_{21}] & g_2 \cdot [k\Lambda'_{22} + (1-k)\cdot\Lambda_{22}] & \cdots & g_2 \cdot [k\Lambda'_{2m} + (1-k)\cdot\Lambda_{2m}] \\ \vdots & \vdots & \cdots & \vdots \\ g_n \cdot [k\Lambda'_{n1} + (1-k)\cdot\Lambda_{n1}] & g_n \cdot [k\Lambda'_{n2} + (1-k)\cdot\Lambda_{n2}] & \cdots & g_n \cdot [k\Lambda'_{nm} + (1-k)\cdot\Lambda_{nm}] \end{array} \right) \end{array}$$

The likelihood matrix Λ_k for the experiment $\mathcal{E}_k$ is obtained by deleting all the g_i terms. We shall abbreviate the experiment's m signal probabilities

$$\sum_{i=1}^n g_i \cdot [k\Lambda'_{i1} + (1-k)\cdot\Lambda_{i1}], \cdots, \sum_{i=1}^n g_i \cdot \left[k\Lambda'_{ij} + (1-k)\cdot\Lambda_{ij}\right], \cdots, \sum_{i=1}^n g_i \cdot [k\Lambda'_{im} + (1-k)\cdot\Lambda_{im}]$$

as

$$w_1(k), \ldots, w_j(k), \ldots, w_m(k).$$

Note that w_j is linear in k. Given the signal $y^k_j \in Y_k$, the posterior probability of the state θ_i is

$$\frac{u_{ij}(k)}{w_j(k)},$$

where

$$u_{ij}(k) \equiv g_i \cdot [k\Lambda'_{ij} + (1-k)\cdot\Lambda_{ij}].$$

Note also that u_{ij} is linear in k, and so is the vector

$$\vec{u}_j(k) = (u_{1j}(k), \ldots, u_{nj}(k)).$$

For each signal y^k_j in the experiment $\mathcal{E}_k$'s signal set Y_k, let $F_{y^k_j}$ denote the cumulative distribution function defined by the n posterior state probabilities $\dfrac{u_{1j}(k)}{w_j(k)}, \ldots, \dfrac{u_{nj}(k)}{w_j(k)}$. When the signal belonging to Y_k has the generic symbol y, and a value of k is understood, then the symbol for the associated posterior distribution function will simply be F_y.

Let the symbol W_k denote the (marginal) cumulative distribution function defined by the signal probabilities $w_1(k), \ldots, w_m(k)$. Using the terminology of our general introductory discussion in 1.4, the information structure associated with the experiment $\mathcal{E}_k$ and the prior G is the triple

$$I_k = \langle Y_k, \{F_y\}_{y \in Y_k}, W_k \rangle.$$

2.3.2 Results that hold with no further restrictions on the two anchor experiments.

The signal y_j^k defines posterior probabilities on the n states $\theta_1, \ldots, \theta_n$, namely the vector $\frac{\vec{u}_j(k)}{w_j(k)}$. To study the typical posterior, we may delete the subscript j (which identifies a particular signal) and we may write the typical vector as $\frac{\vec{u}(k)}{w(k)}$.

Now suppose we are given a real-valued function χ on the probability simplex

$$\Delta(\Theta) = \{(z_1, \ldots, z_n) : \ z_i \geq 0, i \in \{1, \ldots, n\}; z_1 + \cdots + z_n = 1\}.$$

The vector $\frac{\vec{u}(k)}{w(k)}$ belongs to $\Delta(\Theta)$. We will be interested in the expected value of $\chi\left(\frac{\vec{u}(k)}{w(k)}\right)$, where the expectation is taken over the m possible vectors $\frac{\vec{u}(k)}{w(k)}$, each associated with one of the m signals of the structure I_k. We let $\eta_\chi(k)$ denote that expected value.

Without imposing any further conditions on the two anchor structures, we can prove the following theorem.

Theorem B1

If χ is convex (concave), then η_χ is convex (concave).

Proof of Theorem B1:

Suppose χ is convex.

The expected value $\eta_\chi(k)$ is a sum of m terms, each having the form $w(k) \cdot \chi\left(\frac{\vec{u}(k)}{w(k)}\right)$, where $w(k) > 0$ and both $\vec{u}$ and w are linear in k. So it suffices to prove that $w(k) \cdot \chi\left(\frac{\vec{u}(k)}{w(k)}\right)$ is convex in k.

Consider $k_1, k_2 \in [0, 1]$, $\alpha \in [0, 1]$, and $k = \alpha k_1 + (1 - \alpha) \cdot k_2$. We have to show that

$$\chi\left(\frac{\vec{u}(k)}{w(k)}\right)\cdot w(k) \leq \alpha w(k_1)\cdot\chi\left(\frac{\vec{u}(k_1)}{w(k_1)}\right) + (1-\alpha)\cdot w(k_2)\cdot\chi\left(\frac{\vec{u}(k_2)}{w(k_2)}\right) \tag{1}$$

The linearity of $\vec{u}$ and w imply that

$$\vec{u}(k) = \alpha\vec{u}(k_1) + (1-\alpha)\cdot\vec{u}(k_2),\ \ w(k) = \alpha w(k_1) + (1-\alpha)\cdot w(k_2).$$

Hence

$$\begin{aligned}
\chi\left(\frac{\vec{u}(k)}{w(k)}\right)\cdot w(k) &= \chi\left(\frac{\alpha\vec{u}(k_1)+(1-\alpha)\cdot\vec{u}(k_2)}{\alpha w(k_1)+(1-\alpha)\cdot w(k_2)}\right)\cdot(\alpha w(k_1)+(1-\alpha)\cdot w(k_2)) \\
&= \chi\left(\frac{\alpha w(k_1)}{\alpha w(k_1)+(1-\alpha)\cdot w(k_2)}\cdot\frac{\vec{u}(k_1)}{w(k_1)} + \frac{(1-\alpha)\cdot w(k_2)}{\alpha w(k_1)+(1-\alpha)\cdot w(k_2)}\cdot\frac{\vec{u}(k_2)}{w(k_2)}\right) \\
&\qquad \cdot(\alpha w(k_1)+(1-\alpha)\cdot w(k_2)) \\
&\leq \left[\frac{\alpha w(k_1)}{\alpha w(k_1)+(1-\alpha)\cdot w(k_2)}\cdot\chi\left(\frac{\vec{u}(k_1)}{w(k_1)}\right) + \frac{(1-\alpha)\cdot w(k_2)}{\alpha w(k_1)+(1-\alpha)\cdot w(k_2)}\cdot\chi\left(\frac{\vec{u}(k_2)}{w(k_2)}\right)\right] \\
&\qquad \cdot(\alpha w(k_1)+(1-\alpha)\cdot w(k_2)) \\
&= \alpha w(k_1)\cdot\chi\left(\frac{\vec{u}(k_1)}{w(k_1)}\right) + (1-\alpha)\cdot w(k_2)\cdot\chi\left(\frac{\vec{u}(k_2)}{w(k_2)}\right).
\end{aligned}$$

The inequality which comes after the first two equalities follows from (i) the convexity of χ, (ii) the fact that $w(k) > 0$ (which implies that $\lambda w(k_1) + (1-\lambda)\cdot w(k_2) > 0$), and (iii) the fact that

$$\frac{\lambda w(k_1)}{\lambda w(k_1)+(1-\lambda)\cdot w(k_2)} \text{ and } \frac{(1-\lambda)\cdot w(k_2)}{\lambda w(k_1)+(1-\lambda)\cdot w(k_2)}$$

are nonnegative and sum to one. So equation (1) is established.

There is an analogous argument for the case where χ is concave. That completes the proof. □

We shall now use the definition of *regular* in Section 1.7 above, as well as the notation introduced there. Recall that for any distribution F on Θ, the symbol $\tilde{V}_u(F)$ denotes the value of F (the expected payoff when a best quantity is used) for the payoff function u. For a structure I and the payoff function u, the symbol $Q_u(I)$ denotes the expected value of the largest best quantity, over all the posterior distributions defined by the signals of the structure I, and the symbol $V_u(I)$ denotes the expected value of $\tilde{V}_u$. We let $Q_u(k)$ and $V_u(k)$ be abbreviations for $Q_u(I_k)$ and $V_u(I_k)$, respectively, but we omit the subscript when u is understood.

Note that the value function $\tilde{V}_u$ is convex on $\Delta(\Theta)$.[13]

We now apply Theorem B1 to obtain Theorem B2, which again concerns the structures I_k. Like Theorem B1, it imposes no further conditions on the two anchor structures.

Theorem B2

Consider a payoff function u which is regular for I_k for all $k \in [0, 1]$.

(1) The value $V_u(k)$ is convex in k on $[0, 1]$.

(2) Consider the best-quantity function $\hat{q}(\ \cdot\ ; u)$. If $\hat{q}(\ \cdot\ ; u)$ is convex (concave) on $\Delta(\Theta)$, the set of possible distributions on $\Theta = \{(\theta_1, \ldots, \theta_n)\}$, then $Q(\ \cdot\ )$ is convex (concave) on $[0, 1]$.

To establish further results, we shall use the following Lemma concerning convex (concave) functions on $[0, 1]$.

Lemma A

If $f : [0, 1] \to I\!R$ is convex (concave) and $f(k) \geq f(0)$ $(f(k) \leq f(0))$ for all $k \in [0, 1]$, then $f(k_1) \geq f(k_2)$ $(f(k_1) \leq f(k_2))$ whenever $k_1 > k_2 \geq 0$.

Proof of Lemma A:

If $k_2 = 0$, then $f(k_1) \geq f(k_2)$ by assumption. If $k_1 > k_2 > 0$, then the convexity of f implies that $\frac{f(k_1)-f(k_2)}{k_1-k_2} \geq \frac{f(k_1)-f(0)}{k_1-0}$. Since the second fraction is nonnegative and $k_1 - k_2 > 0$, we obtain $f(k_1) \geq f(k_2)$. There is an analogous proof for the case of a concave f. □

Lemma A and Theorem B2 immediately yield the following theorem, which, once again, does not impose further conditions on the anchor structures.

Theorem B3

Consider any payoff function u which is regular for I_k for all $k \in [0, 1]$.

[13] Consider two cumulative distribution functions on Θ, say L and M, and consider $\lambda \in [0, 1]$. Suppose that q^* is a maximizer of $E_{\lambda L+(1-\lambda)\cdot M}\ u(q, \theta)$. Let $\pi(F; u)$ denote $\max_q E_F\ u(q, \theta)$. We have

$$\begin{aligned}\pi(\lambda L + (1-\lambda)\cdot M; u) &= \max_q\ E_{\lambda L+(1-\lambda)\cdot M}\ u(q, \theta) = E_{\lambda L+(1-\lambda)\cdot M}\ u(q^*, \theta) \\ &= \lambda E_L\ u(q^*, \theta) + (1-\lambda)\cdot E_M\ u(q^*, \theta) \\ &\leq \lambda[\max_q\ E_L\ u(q, \theta)] + (1-\lambda)\cdot[\max_q\ E_M\ u(q, \theta)] \\ &= \lambda\pi(L; u) + (1-\lambda)\cdot\pi(M; u).\end{aligned}$$

(1) Suppose

$$(+) \qquad V_u(k) \geq V_u(0) \text{ for all } k \in [0,1].$$

Then we have $V_u(k') \geq V_u(k)$ whenever $1 \geq k' > k \geq 0$.

(2) If the best-quantity function $\hat{q}(\ \cdot\ ; u)$ is convex on $\Delta(\Theta)$ and $Q_u(k) \geq Q_u(0)$ for all $k \in [0,1]$, then $Q_u(k') \geq Q_u(k)$ whenever $1 \geq k' > k \geq 0$.

(3) If $\hat{q}(\ \cdot\ ; u)$ is concave on $\Delta(\Theta)$, and $Q_u(k) \leq Q_u(0)$ for all $k \in [0,1]$, then $Q_u(k') \leq Q_u(k)$ whenever $1 \geq k' > k \geq 0$.

Letting k be our index of effort, the theorem says, informally: (1) if, for a given payoff function, every structure I_k is at least as valuable as the anchor structure I_0, then every structure is at least as valuable as any structure with lower effort; (2) if best quantity is convex in the posterior and average best quantity is never less than the null structure's average best quantity, then the average best quantity is at least as large for a given structure as for a structure with lower effort (Complements); (3) if best quantity is concave in the posterior and average best quantity is never more than the null structure's average best quantity, then the average best quantity is not larger for a given structure than for a structure with lower effort (Substitutes).

We shall see in the next section that if we now specify that the anchor structure I_0 provides no information at all, then we obtain the results of Theorem B3. *without* requiring that for all k, we have $V_u(k) \geq V_u(0)$ and without requiring either $Q_u(k) \geq Q_u(0)$ or $Q_u(k) \leq Q_u(0)$. Those requirements are automatically met.

2.3.3 Requiring the anchor structure I_0 to be a null structure.

If we do not impose a further condition on the anchor structure I_0, we cannot rule out violation of condition (+) in Theorem B3 for some values of the index k. In spite of the value convexity established in Theorem B2, value may be decreasing at some k, even if $V_u(1) > V_u(0)$. It may be that as k increases above zero, V_u first drops below $V_u(0)$ and then rises until it reaches $V_u(1) > V_u(0)$, while still being convex in k on $[0,1]$. We shall see, however, that if the anchor structure I_0 has the *null property*, then when k rises:

- Value never drops when k rises.
- Average best quantity never drops when k rises if best quantity is convex in the posterior.
- average best quantity never rises when k rises if best quantity is concave in the posterior.

In general, for a given joint probability distribution on the state/signal set $\Theta \times Y$, where G is the prior cumulative probability distribution on the states, the structure $I = \langle Y, \{F_y\}_{y \in Y}, W \rangle$

has the null propertys if, for each signal $y \in Y$, we have $F_y = G$, i.e., the signal is useless since the posterior equals the prior. Equivalently, the structure has the null property if y and θ are independently distributed.

In our finite setting, with m states and n signals, a structure has the null property if its n-by-m likelihood matrix Λ is as follows:

$$\Lambda = \begin{pmatrix} q_1 & q_2 & \cdots & q_m \\ q_1 & q_2 & \cdots & q_m \\ \vdots & \vdots & \ddots & \vdots \\ q_1 & q_2 & \cdots & q_m \end{pmatrix},$$

where $q_j > 0$ for all j and $\sum_{j=1}^{n} q_j = 1$.

It is easy to see, using the terminology of our finite Blackwell theorem, that every information structure is at least as informative as a null structure. That is to say, for any regular payoff function u and any prior, the maximal value of $E_{F_y}\ u(q,\theta)$, averaged over all of a structure's signals y, is at least as large for a non-null structure as for a null structure. Moreover, it is not the case that a null structure is at least as informative as a non-null structure: one can always find a prior and a payoff function u such that the maximal value of $E_{F_y}\ u(q,\theta)$, averaged over a structure's signals is higher for the non-null structure than for the null structure.

To see what may happen if we permit the anchor structure I_0 to be a non-null structure, consider the following example.

Let $m = n = 2$ and let the prior probability of each state be $\frac{1}{2}$. Let the likelihood matrix for I_0 be $\Lambda_0 = \begin{pmatrix} \epsilon & 1-\epsilon \\ 1-\epsilon & \epsilon \end{pmatrix}$, with $0 < \epsilon < \frac{1}{2}$. Thus I_0 does not have the null property. Let the likelihood matrix for I_1 be the perfect-information matrix $\Lambda_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Consider $k^* = \frac{\frac{1}{2}-\epsilon}{1-\epsilon}$. We have $0 < k^* < 1$. But we also have

$$k^*\Lambda_1 + (1-k^*)\cdot\Lambda_0 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

So the structure I_{k^*} has the null property even though I_0 does not. It is not true that I_{k^*} is at least as informative as I_0.

We now eliminate such situations by requiring the anchor structure I_0 to be null. We impose no restrictions on the anchor structure I_1. Any mixed structure I_k, with $k \in [0,1]$ is at least as informative as the null structure I_0. So Condition (+) in Theorem B3 is satisfied. Hence for any regular payoff function u, we have $V_u(k') \geq V_u(k)$ whenever $1 \geq k' > k \geq 0$. The other two results of Theorem B3 hold as well.

To establish these claims we first prove another Lemma, which exploits the assumption that the structure I_0 has the null property. The lemma tells us that an increase in k (weakly) increases (decreases) the expected value (over all signals) of *any* convex (concave) function of the posteriors.

Lemma B

Suppose there are $n > 1$ states $\theta_1, \ldots, \theta_n \in \Theta$, with prior probability vector $\vec{g} = (g_1, \ldots, g_n)$. For every $k \in [0,1]$, consider the structure I_k with signal set $Y_k = \{y_1(k), \ldots, y_m(k)\}$ and marginal probability vector $\vec{w}(k) = (w_1(k), \ldots, w_m(k))$. Let $\vec{t}_j(k)$ denote the posterior state-probability vector given the signal $y_j(k)$. Let ϕ be a function from $\Delta(\Theta)$ to the reals. Suppose that I_0 has the null property. Then the following holds:

if ϕ is convex (concave), then

$$E_{y \in Y_{k'}} \ \phi(\vec{t}_j(k')) \geq \ (\leq) \ E_{y \in Y_k} \ \phi(\vec{t}_j(k))$$

whenever $1 \geq k' > k \geq 0$.

Proof of Lemma B:

Since I_0 has the null property, we have

$$\vec{t}_j(0) = \vec{g} \text{ for all } j \in \{1, \ldots, m\}.$$

Suppose ϕ is convex. We have to show that

$$\sum_{j=1}^{m} \ \left[w_j(k') \cdot \phi(\vec{t}_j(k')) \right] \geq \sum_{j=1}^{m} \ \left[w_j(k) \cdot \phi(\vec{t}_j(k)) \right].$$

whenever $1 \geq k' > k \geq 0$. First we have$\sum_{j=1}^{m} \ w_j(k) \cdot \vec{t}_j(k) = \vec{g}$ for all $k \in [0,1]$ (the average of the m posteriors equals the prior). Suppose ϕ is convex. Using Jensen's inequality, we have the following for every $k \in [0,1]$:

$$\sum_{j=1}^{m} \ w_j(k) \cdot \phi(\vec{t}_j(k)) \geq \phi \left(\sum_{j=1}^{m} \ w_j(k) \cdot \vec{t}_j(k) \right) = \phi(\vec{g}) = \phi \left(\sum_{j=1}^{m} \ w_j(k) \cdot \vec{t}_j(0) \right).$$

Now apply Lemma A, letting $f(k)$ be $\sum_{j=1}^{m} \ w_j(k) \cdot \phi(\vec{t}_j(k))$. By Theorem B1, f is convex on $[0,1]$, since ϕ is convex. Since we have just shown that $f(k) \geq f(0)$ for all $k \in [0,1]$, we conclude, using Lemma A, that (as claimed)

$$\sum_{j=1}^{m} \ [w_j(k') \cdot \phi(\vec{u}_j(k'))] \geq \sum_{j=1}^{m} \ [w_j(k) \cdot \phi(\vec{u}_j(k))]$$

whenever $1 \geq k' > k \geq 0$.

The proof for concave ϕ is analogous. □

Let us continue to assume that I_0 has the null property. Let us apply Lemma B, together with Theorem B3, to the convex value function V_u. Doing so, we conclude that

$$V_u(k') \geq V_u(k) \text{ whenever } 1 \geq k' > k \geq 0.$$

Consider, moreover, the best quantity $\hat{q}(F_y; u)$ given a signal y. Lemma B and Theorem B3 tell us that if the best-quantity function $\hat{q}(\,\cdot\,; u)$ is convex (concave), then $Q_u(I_{k'}) \geq Q_u(I_k)$ whenever $1 \geq k' > k \geq 0$.

To summarize, we have established the following theorem.

Theorem B4

Suppose that the anchor structure I_0 has the null property and that u is a regular payoff function for all the posteriors of every structure I_k, i.e., for every $k \in [0,1]$ and every y in $\{y_1(k), \ldots, y_m(k)\}$, there is a largest maximizer of $E_{F_k^y}\; u(q, \theta))$. Then

- The value $V_u(k)$ does not drop when k rises.
- If the best quantity $\hat{q}(\;\cdot\;; u)$ is convex (concave) on $\Delta(\Theta)$, then the average best quantity $Q_u(k)$ does not fall (does not rise) when k rises.

2.3.4 Following a different path to the statements in Theorem B4: calculating a garbling matrix and applying the First Blackwell Theorem.

We now reach the conclusion of Theorem B4 in a different way. We show that the family of structures $\{I_k\}_{k \in [0,1]}$ is, in fact, a Blackwell family. We do so by considering any pair (k', k) with $1 \geq k' > k \geq 0$, and explicitly exhibiting an m-by-m row-stochastic garbling matrix $B_{k'k}$ such that $\Lambda_k = \Lambda_{k'} \cdot B_{k'k}$. We shall find that the garbling matrix is a linear combination of the m-by-m identity matrix and an m-by-m matrix in which all rows are the same. The repeated row has nonnegative elements that sum to one. The first matrix has weight $\frac{k}{k'}$, which falls when k' rises (i.e., when the structure $I_{k'}$ "improves"), while the second matrix has weight $1 - \frac{k}{k'}$. Roughly speaking, the more we improve a structure relative to a given I_k, the more "severe" garbling we need in order to obtain I_k from the improved structure.

In stating the theorem, we let the symbol $\mathbf{e}_t$ denote the t-by-1 matrix of 1s. i.e. $\mathbf{e}_t = (1, 1, \cdots, 1)'$. Then we may write the n-by-m likelihood matrix Λ_0 for the null structure I_0 as $\mathbf{e}_n \cdot \mathbf{q}$, where $\mathbf{q}$ is the row vector $(q_1, q_2, \ldots, q_m)$.

Theorem B5

Consider k, k' with $0 \leq k < k' \leq 1$ and the structures $I_k, I_{k'}$, whose respective n-by-m likelihood matrices are $\Lambda_k = k\Lambda_1 + (1-k) \cdot \Lambda_0$ and $\Lambda_{k'} = k' \cdot \Lambda_1 + (1-k') \cdot \Lambda_0$. Suppose that the structure I_0 has the null property, i.e., $\Lambda_0 = \mathbf{e}_n \cdot \mathbf{q}$, where $\mathbf{q}$ is a row vector $(q_1, \ldots, q_m)$ with nonnegative elements summing to one. Then I_k is a garbling of $I_{k'}$ (and hence $I_{k'}$ is at least as informative as I_k). Specifically, the m-by-m matrix

$$B_{kk'} = \frac{k}{k'}\mathbf{H}_m + \left(1 - \frac{k}{k'}\right)\mathbf{e}_m \cdot \mathbf{q},$$

where $\mathbf{H}_m$ denotes the m-by-m identity matrix, is row-stochastic and satisfies

$$\Lambda_k = \Lambda_{k'} \cdot B_{kk'}.$$

2.3.5 Some remarks on the family of structures $\{I_k\}_{k\in[0,1]}$ considered in the preceding theorems.

Remark 2.3.5.1 We now have two ways of showing that if the anchor structure I_0 has the null property, then (1) the structure $I_{k'}$ is at least as valuable to any Producer as the structure I_k whenever $k' > k$, and (2) we have Complements (Substitutes) if best quantity is a convex (concave) function of the posterior. The first way is based on Theorems B1-B3 and does not use the Blackwell theorems at all. The second way exhibits the garbling matrix that appears in Statement (b) of the First Blackwell Theorem and then uses the fact that statement (b) implies statements (a) and (c). Statement (a) tells us that $I_{k'}$ is at least as informative as I_k. Statement (c) tells us that if best quantity is convex (concave) in the posterior, then average best quantity (over all signals) is not smaller (not larger) for $I_{k'}$ than for I_k.

Remark 2.3.5.2 Even though Theorems B1-B4 are not needed to get results about value or about Complements/Substitutes (since we can get them from the garbling exhibited in Theorem B5 and the First Blackwell Theorem), Theorems B1-B4 have some separate and interesting implications. They provide insight into the Producer's "best" choice of IG effort when the Producer has to pay for the effort. They also allow us to compare the "strength" of the Complements (Substitutes) effects for small k with the strength for large k.

Suppose the Producer has to pay Pk dollars for an information structure with index k, where $P > 0$ is a constant, and seeks to maximize $V_u(k) - Pk$. Since V_u is convex in k, so is $V_u(k) - Pk$. That implies that the maximum of $V_u(k) - Pk$ on $[0, 1]$ occurs at $k = 0$ or at $k = 1$. If the Producer's choice of k is going to be in the *interior* of $[0, 1]$, then *the cost of k has to rise faster than V_u does.*

Now for a posterior F and a regular payoff function u, consider the Producer's largest best quantity, i.e., $\hat{q}(F; u)$, the largest maximizer of $E_F\, u(q, \theta)$. Again let $Q_u(k)$ denote $E_{y\in\{y_1^k,\ldots,y_m^k\}}\, \hat{q}(F^y; u)$.

Suppose $\hat{q}(\ \cdot\ ; u)$ is convex on $\Delta(\Theta)$. Then Theorem B2 tells us that Q_u is convex in k on $[0, 1]$. If we assume that I_0 has the null property, then Theorem B4 tells us that we have Complements, i.e. Q_u is nondecreasing in k. Assume that $Q_u(1) > Q_u(0)$. Then, since Q_u is convex in k , it rises at least as fast as k rises. Roughly speaking, *as the informational effort k rises, the Complements effect of a small increment in effort becomes stronger.*

We have an analogous statement concerning the Substitutes effect when $\hat{q}(\ \cdot\ ; u)$ is concave. Assume that $Q_u(1) < Q_u(0)$. Then, since Q_u is concave in k, it falls at least as fast as k rises. *As the informational effort k rises, the Substitutes effect of a small increment in effort becomes stronger.*

Remark 2.3.5.3 The special case $m = n$ is of particular interest. An example of such an IG is the one-point forecaster discussed in the first preview in Section 1.2 above.[14]

Let the IG's effort be measured by the parameter x, where $x \in [\frac{1}{n}, 1]$. At effort x, the IG's n-by-n likelihood matrix takes the following simple form:

$$\tilde{\Lambda}_x = \begin{pmatrix} x & \frac{1-x}{n-1} & \cdots & \frac{1-x}{n-1} & \frac{1-x}{n-1} \\ \frac{1-x}{n-1} & x & \cdots & \frac{1-x}{n-1} & \frac{1-x}{n-1} \\ \frac{1-x}{n-1} & \frac{1-x}{n-1} & x & \frac{1-x}{n-1} & \frac{1-x}{n-1} \\ \vdots & \vdots & \cdots & \ddots & \vdots \\ \frac{1-x}{n-1} & \frac{1-x}{n-1} & \cdots & \frac{1-x}{n-1} & x \end{pmatrix}.$$

At $x = \frac{1}{n}$, the smallest possible effort, all likelihoods are $\frac{1}{n}$, the likelihood matrix has the null property, and the signals are useless. At $x = 1$, the largest possible effort, the off-diagonal likelihoods are zero, the diagonal likelihoods are one, and each signal is a perfect predictor of the associated state. Replace the variable x with the variable $\tilde{k}(x) = \dfrac{nx - 1}{n - 1}$, which is linear and increasing in x, equals zero when $x = \frac{1}{n}$, and equals one when $x = 1$. Since the likelihood matrix has the null property when $\tilde{k}(x) = 0$, we can apply Theorem B4. Theorem B4 tells us that whenever k increases, value does not drop, or — in the Blackwell terminology — the structure $I_{k'}$ is at least as informative as the structure I_k if $k' > k$. Moreover, Theorem B4 and the First Blackwell Theorem assure us that whenever $\frac{1}{n} \leq x < x' \leq 1$ — which is equivalent to $0 \leq \tilde{k}(x) < \tilde{k}(x') \leq 1$ — the matrix $B_{\tilde{k}(x'),\tilde{k}(x)}$ is a garbling matrix: it is row-stochastic and $\tilde{\Lambda}_x = \tilde{\Lambda}_{x'} \cdot B_{\tilde{k}(x'),\tilde{k}(x)}$. So the family of structures indexed by x is indeed a Blackwell family and higher x — which we may interpret as improving the reliability of the forecasts — indeed benefits the Producer. Moreover, we have Complements (Substitutes) if the Producer's best quantity is

[14] The term "all or nothing" has sometimes been used to describe a one-point forecaster when the signal set is the same as the state set and all states have equal prior probability (or equal prior density). Given a signal y the true state is y with probability x. With probability $1 - x$ all states are equally probable, which the prior already told us. Two papers that consider all-or-nothing structures, but in very different settings, are Johnson and Myatt (2006), and Rajan and Saouma (2006).

convex (concave) in the posterior: higher x means that the average value of the best quantity, over the n possible forecasts, rises or stays the same (falls or stays the same).

Remark 2.3.5.4. There are many Blackwell families of information structures where the "corner" phenomenon described in Remark 2.3.5.2 does *not* arise. That is true, in particular, if the structures are indexed by $k \in \{1, 2, 3, \dots\}$ and for every regular payoff function u, perfect information provides an upper bound to expected payoff. Suppose the family $\{I_k\}_{k \in \{1,2,3,\dots\}}$, where $I_k = \langle Y^k, \{F_y^k\}_{y \in Y^k} \rangle$, is a Blackwell family with respect to k. That is to say, for every regular u, we have

$$V_u(k') \geq V_u(k) \text{ whenever } k' > k,$$

where $V_u(k)$ again denotes the average, over all signals $y \in Y^k$, of the highest attainable value of $E_{F_y^k}\ u(q, \theta)$. Let $u^*(\theta)$ denote the perfect-information expected payoff for the state θ, i.e. $u^*(\theta) = \max_q u(q(\theta), \theta)$ Define

$$B_u \equiv E_\theta\ u^*(\theta).$$

Then

$$(\dagger) \qquad V_u(k) \leq B_u \text{ for all } k \in \{1, 2, 3, \dots, \}$$

Now suppose again that the cost of the index k is $P \cdot k$, where $P > 0$. Then if $V_u(k) - Pk > 0$ for some k, there exists a "best" value of k, i.e.,[15]

$$(\dagger\dagger) \qquad \text{there exists } k^* \text{ such that } V_u(k) - Pk \leq V_u(k^*) - Pk^* \text{ for all } k.$$

If we now truncate the family, so that the set of possible indices is, say, $\{1, 2, \dots, \bar{k}\}$, then k^* is an interior "best" index if $k^* < \bar{k}$. An example of a Blackwell family indexed by $k \in \{1, 2, 3, \dots\}$, in which perfect information provides an upper bound, is a family in which the structure I_k partitions the state set θ into k sets and the partitioning for k is a refinement of the partitioning for $k - 1$.

Remark 2.3.5.5. Consider the joint probability matrix in section 2.3.1. One can extend the matrix so that the rows correspond to an infinite but countable state set $\Theta = \{\theta_1, \theta_2, \theta_3, \dots, \}$ and columns correspond to an infinite but countable signal set $\{y_1^k, y_2^k, y_3^k, \dots\}$. (An example is the case where θ has a Poisson distribution). There is again a prior on Θ. For $k = 0$, the posterior given any signal again equals the prior. For $k = 1$, that is not the case. The likelihood matrix Λ_k is a countable-row and countable-column version of the matrix Λ_k discussed in section 2.3.1. The definition of garbling extends to such matrices. We again obtain versions of Theorems B1-B4, using essentially the same proofs as in the preceding sections. We omit the details.

[15] Since, by assumption, $V_u(k) - Pk > 0$ for some k, it follows that $(\dagger\dagger)$ holds if there exists $\tilde{k}$ such that

$$\text{for all } k > \tilde{k} \text{ we have } V_u(k) - Pk < 0.$$

But the existence of such a $\tilde{k}$ is obvious since $V_u(k) - Pk \leq B_u - Pk < 0$ when $k > \frac{B_u}{P}$.

2.4 Counterparts of the previous results when the state and signal sets are continua.

We now consider a non-finite version of the collection $\{I_k\}$ of finite structures. For the finite collection, consider again the case $m = n$, which we called the case of a "one-point forecaster" in Remark 2.3.5.3. In the analog we now study, both the state set Θ and a structure's signal set are continua. Each signal predicts a unique state and for each state there is a unique signal that predicts it. Every structure is again a linear combination of two anchor structures I_0 and I_k, with weight $k \in [0, 1]$.

We shall obtain Theorems $\tilde{B}1$-$\tilde{B}5$. They are non-finite counterparts of our finite theorems B1-B5. Theorem $\tilde{B}1$ again tells us that if a function χ on the posteriors is convex (concave), then η_χ, the expected value of χ over the sisgnals of the k-structures, is convex (concave) on $[0, 1]$. Theorem $\tilde{B}1$ again implies Theorem $\tilde{B}2$, concerning the convexity on $0, 1$ of the value $V_u(k)$ and the convexity/concavity of the best-quantity $Q_u(k)$. Lemma A and theorem $\tilde{B}2$ again imply Theorem $\tilde{B}3$, which does not require the first of the two anchor structures to have the null property and concerns the direction in which Q_u and V_u move when k increases. We do not state Theorem $\tilde{B}3$, since its wording is the same as Theorem $B3$.

Theorem $\tilde{B}4$ again requires the anchor structure I_0 to be null, which again yields sharper results. The theorem again tells us that if u is regular for every I_k, then when k rises, the value function $V_u(k)$ (weakly) rises and the average quantity $Q_u(k)$ (weakly) rises if the best-quantity function is convex and (weakly) falls if the best-quantity function is concave. Theorem $\tilde{B}5$ again tells us that I_k is a garbling of $I_{k'}$ whenever $0 \leq k < k' \leq 1$. The proof explicitly exhibits the required Markov kernel.

So we again have two paths leading to the conclusions of Theorem $\tilde{B}4$. The first path uses only Theorems $\tilde{B}1 - \tilde{B}4$ and does not use a Blackwell theorem. The second path uses only Theorem $\tilde{B}5$, where a Markov kernel is explicitly shown, together with the Fourth Blackwell Theorem.

The comments made in Remark 2.3.5.2 for the finite case apply again in the general case. We again need the cost associated with the index k to rise sufficiently fast if the "best" k is going to be in the interior of $[0, 1]$. We again see that if the best-quantity function is convex, then the Complements effect of a small increment in k becomes stronger as k increases; and if the best-quantity function is concave, then the Substitutes effect of a small increment in k becomes stronger as k increases.

Recall the discussion of the general case in Section 1.4. The prior beliefs about the states are expressed by a probability measure space

$$(\Theta, \mathcal{T}, G).$$

An experiment specifies, for each state θ, a likelihood measure space

$$L_\theta = (Y, \mathcal{Y}, \lambda_\theta).$$

An experiment on Θ is a collection $\mathcal{E} = \{L_\theta\}_{\theta \in \Theta}$. Then if we fix $\mathcal{T}$ and G, we have an associated information structure, composed of a signal set, a posterior state-probability measure on $\mathcal{T}$ for each signal, and a marginal signal-probability measure.

The experiment $\mathcal{E}_k$ now specifies, for each state θ, a likelihood measure space

$$L_\theta^k = (Y, \mathcal{Y}, \lambda_\theta^k).$$

Every experiment $\mathcal{E}_k$ has the same signal set Y and the same σ-algebra $\mathcal{Y}$. But as we move from one experiment to another (with G and $\mathcal{T}$ fixed), the posterior state measures on $\mathcal{T}$ change and so does the likelihood measure space for a fixed θ. The information structure associated with $\mathcal{E}_k$ (with G and $\mathcal{T}$ fixed) is

$$I_k = \langle Y, \{F_y^k\}_{y \in Y}, W_k \rangle,$$

where F_y^k is the posterior state measure when the signal is y and W_k is the marginal signal-probability measure on $\mathcal{Y}$.

We shall assume that G has a density function g, where $g(\theta) > 0$ for all $\theta \in \Theta$. We also assume that for each θ, the likelihood measure λ_θ^k has a positive-valued density function. For each $y \in Y$, the likelihood density will be denoted

$$\lambda^k(y \mid \theta).$$

Finally, we have, for each k, a marginal signal density h^k defined by

$$h^k(y) = \int_\Theta \lambda^k(y|\theta)g(\theta)d\theta.$$

Note that $h^k(y) > 0$ for all y, since, by assumption, $\lambda^k(y|\theta) > 0$ and $g(\theta) > 0$.

For the structure I_k and for a given signal y, we have a posterior measure F_y^k on $\mathcal{T}$. That measure has a density. For a given θ, the density is denoted $f^k(\theta|y)$. We have

$$f^k(\theta|y) = \frac{\lambda^k(y|\theta) \cdot g(\theta)}{h^k(y)}.$$

For every I_k we have the mixture property

$$\lambda^k(y|\theta) = (1-k) \cdot \lambda^0(y|\theta) + k \cdot \lambda^1(y|\theta).$$

Note that

$$h^k(y) = \int_\Theta \lambda_k(y|\theta)g(\theta)d\theta = \int_\Theta [(1-k)\cdot\lambda^0(y|\theta) + k\cdot\lambda^1(y|\theta)]$$
$$= (1-k)\cdot h^0(y) + kh^1(y).$$

Now let $\Delta(\Theta)$ denote the set of probability measures on $\mathcal{T}$. We will abuse the "$\chi(\ \cdot\ )$" notation: for a measure μ with density function m, the symbol $\chi(m)$ will be interpreted as $\chi(\mu)$. Since, in particular, the measure F_y^k has density function $f^k(y|\theta)$, we have

$$\eta_\chi(k) \equiv \int_Y h^k(y)\cdot\chi\left(\frac{g(\theta)\cdot\lambda^k(y|\theta)}{h^k(y)}\right)dy.$$

Theorem $\tilde{B}1$ now follows. Its proof repeats the pattern of the proof of Theorem B1.

Theorem $\tilde{\mathbf{B}}$1

Consider the probability measure space $(\Theta, \mathcal{T}, G)$ and the structure family $\{I_k\}_{k\in[0,1]}$, where $I_k = \langle Y, \{F_y^k\}_{y\in Y}, W_k\rangle$. Let G have a positive-valued density function g, and for each k and every $\theta\in\Theta$ let the likelihood measure λ_θ^k have a positive-valued density function. Suppose χ is convex (concave). Then η_χ is convex (concave) on $[0,1]$.

Theorem $\tilde{B}2$, the general counterpart of Theorem B2, now follows, in the same way as before, from Theorem $\tilde{B}1$ and the convexity of the function $\tilde{V}_u$. (Recall that $\tilde{V}_u(F)$ denotes $\max_q(E_F\ U(q,\theta))$).

Again we call a Producer's payoff function $u : \mathbb{R}\times\Theta\to\mathbb{R}$ *regular for the structure* I_k if for each of that structure's possible posteriors there is a largest maximizer of the expected value of u under the posterior. Recall that the symbol $V_u(k)$ denotes the highest attainable expected payoff, averaged over the structure's possible posteriors.

Theorem $\tilde{B}2$

Consider the probability measure space $(\Theta, \mathcal{T}, G)$ and the structure family $\{I_k\}_{k\in[0,1]}$, where $I_k = \langle Y, \{F_y^k\}_{y\in Y}, W_k\rangle$. Let G have a positive-valued density function g, and for each k and every $\theta\in\Theta$ let the likelihood measure λ_θ^k have a positive-valued density function. For any payoff function u which is regular for I_k for all $k\in[0,1]$, the value $V_u(k)$ is convex in k on $[0,1]$.

Lemma A does not involve the set of states and remains unchanged. As before, Lemma A and Theorem $\tilde{B}2$ imply Theorem $\tilde{B}3$, which repeats Theorem B3 except that the state set $\Theta = \{\theta_1,\ldots,\theta_n\}$ is replaced by an arbitrary set Θ.

Next we establish Lemma $\tilde{B}$, a counterpart of Lemma B, which again deals with the case where the anchor structure I_0 has the hull property. In the statement of the lemma, an arbitrary state set Θ and a prior measure G on the σ-algebra $\mathcal{T}$ are given. To say that I_0 has the null property means that for every signal y, we have $F_y^0 = G$, or equivalently

there exists a function $\ell : Y \to I\!R^+$ such that for all $y \in Y$, we have $\lambda^0(y|\theta) = \ell(y)$.

That statement is equivalent, in turn, to the statement[16]

$$h^0(y) = \lambda^0(y|\theta) \text{ for all } \theta \in \Theta.$$

Lemma $\tilde{B}$

Let ϕ be a real-valued function on Δ. Suppose that I_0 has the null property. Then the following holds:

if ϕ is convex (concave), then

$$E_{y\in Y_{k'}}\ \phi(F_y^{k'}) \geq\ (\leq)\ \ E_{y\in Y_k}\ \phi(F_y^k) \text{ whenever } 1 \geq k' > k \geq 0.$$

Proof of Lemma $\tilde{B}$:

We have $\int_Y\ h^k(y) \cdot F_y^k dy = G$. (The average of all the posteriors equals the prior). Suppose ϕ is convex. Then, using Jensen's inequality, we have the following for every $k \in [0,1]$:

$$\int_Y\ h^k(y) \cdot \phi(F_y^k) dy \geq \phi\left(\int_Y\ h^k(y) \cdot F_y^k dy\right) = \phi(G) = \phi\left(\int_Y\ h^0(y) \cdot F_y^0 dy\right).$$

(The last equality follows from the null property of I_0).

Now apply Lemma A, letting $f(k)$ be $\int_Y\ h^k(y) \cdot \phi(F_y^k) dy$. By Theorem $\tilde{B}1$, f is convex on $[0,1]$, since ϕ is convex. Since we have just shown that $f(k) \geq f(0)$ for all $k \in [0,1]$, we conclude, using Lemma A, that (as claimed)

$$E_{y\in Y_{k'}}\ \phi(F_y^{k'}) \geq E_{y\in Y_k}\ \phi(F_y^k)$$

whenever $1 \geq k' > k \geq 0$.

The proof for concave ϕ is analogous. □

[16] We have

$$h^0(y) \ = \ \int_\Theta\ \lambda^0(y|\theta) g(\theta) d\theta = \int_\Theta\ \ell(y) g(\theta) d\theta = \ell(y) \int_\Theta\ g(\theta) d\theta = \ell(y) = \lambda^0(y|\theta).$$

Analogously to the finite-state-set case, Lemma $\tilde{B}$, together with Theorem $\tilde{B}3$, yield an analogue of Theorem B4, namely the following Theorem $\tilde{B}4$, where the state set and th signal set need not be finite.

Theorem $\tilde{B}4$

Consider the probability measure space $(\Theta, \mathcal{T}, G)$ and the structure family $\{I_k\}_{k\in[0,1]}$ where $I_k = \langle Y, \{F_y^k\}_{y\in Y}, W_k\rangle$. Let G have a positive-valued density function g, and for each $k \in [0,1]$ and every $\theta \in \Theta$ let the likelihood measure λ_θ^k have a positive-valued density function. Let I_0 have the null property. Let u be a regular payoff function for all the posteriors of every structure I_k. Then

- The value $V_u(k)$ does not drop when k rises.
- If the best-quantity function $\hat{q}(\ \cdot\ ; u)$ is convex (concave) on Δ, then the average best quantity $Q_u(k)$does not fall (does not rise) when k rises.

Next we impose the null condition on the anchor structure I_0. That means that for this structure

$$h^0(y) = \lambda^0(y|\theta).$$

Next we have Theorem $\tilde{B}5$, a counterpart of Theorem B5. Theorem $\tilde{B}5$ tells us that if the anchor structure I_0 has the null property, then for the structures $I_k, I_{k'}$, with $k' > k$, we have the garbling required in statement (b) of the Third and Fourth Blackwell Theorems. Theorem $\tilde{B}5$ explicitly exhibits the required Markov kernel.

The Fourth Blackwell Theorem tells us that statement (b) implies statement (a). If we now assume (as in Theorem $\tilde{B}4$) that every likelihood measure has a density, then the measure is absolutely continuous, as assumption A1 in the Fourth Blackwell Theorem requires. Then our Theorem $\tilde{B}5$ and the Fourth Blackwell Theorem tell us that whenever $k' > k$, structure $I_{k'}$ is at least as valuable for any Producer as structure I_k. Moreover, the Fourth Blackwell Theorem says that statement (b) implies statement (c). So we can claim that if the Producer's best-quantity function is convex (concave), then moving from I_k to $I_{k'}$, where $k' > k$ cannot raise (cannot diminish) the average best quantity. So we have Complements (Substitutes).

Thus Theorem $\tilde{B}5$ provides our alternative Blackwell-theorem path to the statements in the conclusion of Theorem $\tilde{B}4$.

Theorem $\tilde{B}5$

Suppose that the anchor experiment $\mathcal{E}_0 = (Y, \mathcal{Y}, \{F_y^0\}_{y\in Y})$ has the null property, i,e., there exists a probability measure h^0 on $(Y, \mathcal{Y})$ such that

$$h^0(A) = \lambda_\theta^0(A) \text{ for all states } \theta \in \Theta \text{ and all sets } A \in \mathcal{Y}.$$

Then

$$\mathcal{E}_k \text{ is a garbling of } \mathcal{E}_{k'} \text{ whenever } 1 \geq k' > k \geq 0.$$

2.5 The *erratic IG*: another Blackwell family of finite information structures indexed by $k \in [0,1]$, where k is the probability that the IG uses the "better" of two structures.

Now consider another family of finite structures, also indexed by $k \in [0,1]$. The state set Θ again has n elements. Their probabilities $p_1, \ldots, p_n$ are again given. Let C be a c-element set, and let D be a d-element set. Let Λ be an n-by-d row-stochastic matrix and let Λ^* be an n-by-c row-stochastic matrix. For the structure I_k, with $k \in [0,1]$, the signal set is $C \cup D$, the set of posterior PDF's is $\{F_y^k\}_{y \in C \cup D}$, and the likelihood matrix is the n-by-$(c+d)$ row-stochastic matrix

$$\Lambda_k = (k \cdot \Lambda^*, (1-k) \cdot \Lambda).$$

The interpretation is that the IG will use one of two information structures. He is unsure which one it will be. In a more specific scenario, the IG uses one of two experts. The first expert has signal set C and likelihood matrix Λ^*. The second has signal set D and likelihood matrix Λ. The second expert is always available but the first is only available with probability k. If the first expert is available, he will be used; if not, the second expert is used. More effort by the IG means that he induces the first expert, by an appropriate payment, to be available more frequently and thereby increases k. This is of particular interest if the first expert is more informative than the second.[17]

We will need symbols for the vector of n posterior state probabilities given a signal y If $y \in C$, the symbol $\vec{\pi}^C$ will be used for that vector; if $y \in D$, the symbol $\vec{\pi}^D$ will be used. For $y \in C \cup D$, we define

$$\vec{\pi}(y) \equiv \begin{cases} \vec{\pi}^C(y) & \text{if } y \in C \\ \vec{\pi}^D(y) & \text{if } y \in D. \end{cases}$$

We let the symbol $E_{\vec{\pi}(y)}$ denote expectation when the state-probability vector $\vec{\pi}(y)$ is used to calculate the expectation. When needed we shall add the subscript k, to indicate that we are considering the signals of the structure I_k. We then have $\vec{\pi}_k^C(y), \vec{\pi}_k^D(y), \vec{\pi}_k(y)$.

Now consider a regular payoff function u. For that payoff function, the value of the structure

$$I_k = \langle C \cup D, \{F_y^k\}_{y \in C \cup D} \rangle$$

[17] A model in which decision-makers receive "no additional information" with probability p and useful (but not perfect) information with probability $1-p$ is studied, in a very different context, by Green and Stokey (2003) .

is

$$V_u(k) = E_{y \in C \cup D}\Bigg(E_{F_y^k}\ u\left(\hat{q}(F_y^k; u), \theta\right)\Bigg).$$

(Recall that $\hat{q}(F; u)$ denotes a maximizer of $E_F\ u(q, \theta)$).

We shall slightly abuse the $\hat{q}(F; u)$ notation. We omit the symbol u and we let $\hat{q}(\vec{\pi}(y))$ denote a quantity which maximizes the expected value of u when the state-probability vector $\vec{\pi}(y)$ is used to calculate the expectation. We can then rewrite the value $V_u(k)$ as

$$V_u(k) = E_{y \in C \cup D}\ \Bigg(E_{\vec{\pi}_k(y)}\ u(q^*(\vec{\pi}_k(y)), \theta)\Bigg).$$

Note that the $(c+d)$-by-$(c+d)$ likelihood matrices for the structures I_0, I_1 are, respectively, $\Lambda_0 = (\mathbf{0}_{nd}, \Lambda)$ and $\Lambda_1 = (\Lambda, \mathbf{0}_{nc})$, where the symbol $\mathbf{0}_{rs}$ denotes the r-by-s zero matrix. If $k = 1$, the signals in D do not occur; if $k = 0$, the signals in C do not occur. Thus

$$(+) \qquad V_u(1) = E_{y \in C}\Bigg(E_{\vec{\pi}_1^C(y)}\ u(q^*(\vec{\pi}_1^C(y)), \theta)\Bigg), \quad V_u(0) = E_{y \in D}\Bigg(E_{\vec{\pi}_0^D(y)}\ u(q^*(\vec{\pi}_0^D(y)), \theta)\Bigg).$$

We now obtain a somewhat degenerate analogue of Theorem B1.

Consider any function $\chi : \Delta(\Theta) \rightarrow I\!R$. Let $\eta_\chi(k)$ denote the average, over all the signals y in the structure I_k, of $E_{\vec{\pi}_k(y)}\ \chi(\vec{\pi}_k(y))$.

Theorem B6

For any function $\chi : \Delta \rightarrow I\!R$, the function η_χ is linear in k.

Now let $\hat{\chi}$ denote the function of the structure I_k's posteriors that we considered above, namely the function

$$E_{\vec{\pi}(y)}\ u\left(q^*(\vec{\pi}(y)), \theta\right).$$

In the proof of theorem B5 we establish:

$$\eta_\chi(k) = k \cdot \sum_{y_j \in C}\ [\sum_{i=1}^n \Lambda^*_{ij} \cdot p_i] \cdot \chi(\vec{\pi}^C(y_j)) + (1-k) \cdot \sum_{y_\ell \in D}\ [\sum_{i=1}^n \Lambda_{i\ell} \cdot p_i] \cdot \chi(\vec{\pi}^D(y_\ell)).$$

Note that (+) implies

$$\sum_{y_j \in C}\ \big([\text{ prob. of } y_j] \cdot \hat{\chi}(\vec{\pi}^C(y_j))\big) = \sum_{y_j \in C}\ [\sum_{i=1}^n \Lambda^*_{ij} \cdot p_i] \cdot \hat{\chi}(\vec{\pi}^C(y_j)) = V_u(1)$$

and

$$\sum_{y_\ell \in D} \big([\text{ prob. of } y_\ell] \cdot \hat{\chi}(\vec{\pi}^D(y_j))\big) = \sum_{y_\ell \in D} [\sum_{i=1}^{n} \Lambda_{i\ell} \cdot p_i] \cdot \hat{\chi}(\vec{\pi}^D(y_\ell)) = V_u(0).$$

Since

$$\sum_{y_j \in C} \big([\text{ prob. of } y_j] \cdot \hat{\chi}(\vec{\pi}^C(y_j))\big) = V_u(1)$$

and

$$\sum_{y_\ell \in D} \big([\text{ prob. of } y_\ell] \cdot \hat{\chi}(\vec{\pi}^D(y_\ell))\big) = V_u(0).$$

We conclude that

$$(++) \qquad V_u(k) = k \cdot V_u(1) + (1-k) \cdot V_u(0).$$

Next suppose that I_1 is more informative than I_0, i.e., for every regular u we have $V_u(1) \geq V_u(0)$. Then, in view of (++),

$(+++)$ for every regular u we have $V_u(k') \geq V_u(k)$ whenever $1 \geq k' > k \geq 0$.

(That is the counterpart of Theorem B3). As before, there is an alternative path to the statement (+++). By the Blackwell theorem, (+++) is equivalent to the statement

$(+++++)$ there exists a $(c+d)-\text{by}-(c+d)$ row-stochastic matrix G such that $\Lambda_k = \Lambda_{k'} \cdot G$.

We now exhibit G explicitly. First note that $V_u(1) \geq V_u(0)$ for all regular u (by (+++)). That implies (in view of the zero matrices that enter the likelihood matrices Λ_0, Λ_1) that there exists a c-by-d row-stochastic matrix B such that

$$\Lambda = \Lambda^* \cdot B.$$

The matrix B is used in constructing the garbling matrix G. The matrix G turns out to be as follows:

$$G = \begin{pmatrix} \frac{k}{k'} \cdot \mathbf{H}_c & \big(1 - \frac{k}{k'}\big) \cdot B \\ \mathbf{0}_{dc} & \mathbf{H}_d \end{pmatrix},$$

where, as in Theorem B5, the symbol $\mathbf{H}_\ell$ denotes the ℓ-by-ℓ identity matrix.

It is straightforward to check that the matrix G just defined is row-stochastic. Moreover, we have

$$\begin{aligned} \Lambda_{k'} \cdot G &= (k' \cdot \Lambda^*, (1-k') \cdot \Lambda) \cdot G \\ &= \left(k' \cdot \Lambda^* \cdot \frac{k}{k'} \cdot \mathbf{H}_c + (1-k') \cdot \Lambda \cdot \mathbf{0}_{dc}, k' \cdot \Lambda^* \cdot \left(1 - \frac{k}{k'}\right) \cdot B + (1-k') \cdot \Lambda \cdot \mathbf{H}_d\right) \\ &= \Big(k \cdot \Lambda^*,\ (k' - k + 1 - k') \cdot \Lambda\Big) = (k \cdot \Lambda^*, (1-k) \cdot \Lambda) = \Lambda_k, \end{aligned}$$

as required in (++++). (We obtain the next-to-last equality from the one that precedes it by using the fact that $\Lambda = \Lambda^* \cdot B$).

Note that since $V_u(k)$ is nondecreasing in k, we again have the statement in Remark 3.5.1 about Complements/Substitutes (that statement uses the finite Blackwell theorem). We also have, as in Remark 3.5.2, the "corner" property of a "best" index k when the cost of k is linear in k.

2.6 Sufficient conditions for best quantity to be a convex (concave) function of the posterior.

Consider a Blackwell IG. Suppose we index the IG's possible information structures with an index ν, where the first of two structures is at least as informative as the second if the first has a higher value of ν. The IG expends more effort when ν increases. We have Complements (Substitutes) for a given Producer if the Producer's largest expected-payoff-maximizing quantity is a convex (concave) function on the set of possible posteriors. In the Blackwell applications presented below, in Section 4.1, we establish the required convexity or concavity directly by exploiting special properties of particular payoff functions. But it is natural to seek more general conditions — on the posteriors and the payoff function — implying convexity or concavity. We obtain such conditions in the present section.

As before, the random state of the world is θ , and the set of its possible values is Θ. A σ-algebra $\mathcal{T}$ of subsets of Θ is given and stays the same throughout the discussion. We let $\mathcal{F}$ denote a set of possible posterior probability measures on $\mathcal{T}$. The Producer's quantity is q and its possible values comprise the set Q. The producer's payoff is $u : Q \times \Theta \to I\!R$.

For a fixed $q \in Q$, we shall let the symbol $\breve{u}(q, F)$ denote the expected value of u under the posterior F, i.e.,

$$\breve{u}(q, F) = \int_S u(q, \theta) dF.$$

Now assume that for fixed θ, u is differentiable with respect to q.

Let

$$\eta(q, F) \equiv \frac{\partial}{\partial q} \breve{u}(q, F).$$

Theorem B7 Suppose that

(i) For all $F \in \mathcal{F}$, the equation $\eta(q, F) = 0$ has a unique solution in Q, denoted $q^*(F)$.

(ii) $\eta(q, F) > 0$ for $q < q^*(F)$ and $\eta(q, F) < 0$ for $q > q^*(F)$.

(iii) There exist functions η^*, η^{**} and a strictly increasing function ξ, such that $\eta = \eta^* - \eta^{**}$ and the function $\xi(\eta^*(q,F))$ is convex in *the pair* (q,F), while the function $\xi(\eta^{**}(q,F))$ is concave in *the pair* (q,F).

Then the function $\hat{q}$ is convex.

An analogous proof, which we omit, establishes the following theorem.

Theorem B8 Suppose that

(i) For all $F \in \mathcal{F}$, the equation $\eta(q,F) = 0$ has a unique solution in Q, denoted $q^*(F)$.

(ii) $\eta(q,F) > 0$ for $q < q^*(F)$ and $\eta(q,F) < 0$ for $q > q^*(F)$.

(iii) There exist functions η^*, η^{**} and a strictly increasing function ξ, such that $\eta = \eta^* - \eta^{**}$ and the function $\xi(\eta^*(q,F))$ is concave in *the pair* (q,F), while the function $\xi(\eta^{**}(q,F))$ is convex in *the pair* (q,F).

Then the function $\hat{q}$ is concave.

3 Non-Blackwell information-gatherers.

3.1 Three theorems about an IG who constructs equal-probability partitionings, or equal-length partitionings, of the state set.

First we specify that Θ is an interval $[u,v]$ in $I\!R$, where $u < v$ and we may have $u = -\infty$, or $v = \infty$, or both. There is a prior cumulative probability distribution G on Θ. So before the IG expends effort, the producer's prior probability for the event $u \leq \theta \leq \theta^*$ is $G(\theta^*)$. We shall prove three Theorems, denoted N1,N2,N3, where "N" stands for *Non-Blackwell.* In the first two theorems, the Information-gatherer devotes effort to the partitioning of $[u,v]$ into n intervals that have *equal probability*; in the third theorem, the intervals have *equal length.* More effort means more intervals. In all three theorems, the IG's family of structures is *not* a Blackwell family. The $(n-1)$-interval structure is not a garbling of the n-interval structure. It is straightforward to establish this by exhibiting a Producer payoff function, a prior G, and a poitive integer n such that highest attainable expected payoff is higher for n than for $n+1$. In Theorem N1 and N3, $\Theta = [0,1]$ and the prior G has the form $G(\theta) = \theta^{\frac{1}{a}}$ for some $a > 0$. (The uniform distribution is an example). In Theorem N2, $\Theta = [k,\infty)$, where $k > 0$, and the prior G has the form $G(\theta) = 1 - \left(\frac{k}{\theta}\right)^{\delta}$, where $\delta > 0$. (The Pareto-Levy distribution is an example).

The IG's family of structures is

$$\{I_n \ : n = 1,2,3,\dots\},$$

where

$$I_n = \langle Y^n, \{F_y^n\}_{y \in Y^n}, W_n \rangle.$$

The structure's signal set Y^n contains n signals, denoted $y_1^n, \ldots, y_n^n$. Each of them uniquely identifies one of the partitioning's intervals. We shall let the symbol y_i^n denote both a signal and the interval it identifies.[18]

The theorems will be applied to a Producer whose best quantity, given a signal y, depends only on $E_{F_y}\ \theta$, the mean of the posterior that is associated with that signal. The theorems are concerned with a function $\phi : \Theta \to I\!R$. They describe the effect of more IG effort (higher n) on the expected value, over all signals y, of $\phi\left(E_{F_y}\ \theta\right)$, where F_y is the posterior determined by y. The therems consider the case where ϕ takes the form $\phi(t) = t^m$ and m can be any real number. In some applications $\left(E_{F_y}\ \theta\right)^m$ will be the Producer's best quantity given the signal y, and in other applications it will be the Producer's maximal expected payoff given the signal y. Several applications will be presented in section 4.2.

The theorems state conditions under which the expected value of $\left(E_{F_y}\ \theta\right)^m$, over all the signals y in Y^n, is increasing in n, decreasing in n, and constant with respect to n. In all three theorems, we find that this expected value is strictly increasing in n if $m > 1$ or $m < 0$, strictly decreasing if $0 < m < 1$, and constant if $m = 1$ or $m = 0$.

Theorem N1

Let the state set Θ be $[0, 1]$ and let the prior G on Θ satisfy $G(\theta) = \theta^{\frac{1}{a}}$ for some $a > 0$. Consider the function ϕ, where

$$\phi(t) = t^m,$$

and the sequence

$$A_1, A_2, \ldots, A_n \ldots,$$

where

$$A_n = \frac{1}{n} \sum_{i=1}^{n} \phi(E_{F_{y_i}^n}\ \theta)$$

and y_i identifies the ith interval in the equal-probability intervals

$$[x_0^n, x_1^n], (x_1^n, x_2^n], \ldots, (x_i^n, x_{i+1}^n], \ldots, (x_{n-2}^n, x_{n-1}^n], (x_{n-1}^n, x_n^n].$$

Then for all $a > 0$:

[**1**] the sequence is strictly increasing (i.e., $n' > n \geq 1$ implies $A_{n'} > A_n$) if $m > 1$ or $m < 0$;

[18] The structure I_n is associated with a "noiseless" experiment $\mathcal{E}_n$. In this experiment the likelihood function for every θ assigns probability one to the signal that identifies the interval in which θ lies and probability zero to every other signal.

[**2**] the sequence is strictly decreasing (i.e., $n' > n \geq 1$ implies $A_{n'} < A_n$) if $0 < m < 1$;

[**3**] the sequence is constant if $m = 1$ or $m = 0$.

Theorem N2

Let the state set Θ be $[k, \infty)$, where $k > 0$ and let the prior G on Θ satisfy $G(\theta) = 1 - \left(\frac{k}{\theta}\right)^{\delta}$, with $\delta > 1$. Consider the function ϕ, where

$$\phi(t) = t^m,$$

and the sequence

$$A_1, A_2, \ldots, A_n \ldots,$$

where

$$A_n = \frac{1}{n} \cdot \sum_{i=1}^{n} \phi\left(E_{F^n_{y_i}}\ \theta\right)$$

and y_i identifies the ith interval in the equal-probability intervals

$$[x^n_0, x^n_1], (x^n_1, x^n_2], \ldots, (x^n_i, x^n_{i+1}], \ldots, (x^n_{n-2}, x^n_{n-1}], (x^n_{n-1}, x^n_n].$$

The following statements hold:

[**1**] the sequence is strictly increasing (i.e., $n' > n \geq 1$ implies $A_{n'} > A_n$) if $m > 1$ or $m < 0$;

[**2**] the sequence is strictly decreasing (i.e., $n' > n \geq 1$ implies $A_{n'} < A_n$) if $0 < m < 1$;

[**3**] the sequence is constant if $m = 1$ or $m = 0$.

Theorem N3 concerns equal-length intervals. We may interpret a partitioning into equal-length intervals as "rounding off" the true state θ to a fixed number of decimal places; More intervals correspond to more precision. If the IG chooses $n = 10^d$, then his effort is a once-and-for-all investment which provides him with the capability of correctly identifying the current state to 10^d decimal places, regardless of the frequency with which a particular interval occurs.

THEOREM N3

Let the state set Θ be $[0, 1]$ and let the prior G on Θ satisfy $G(\theta) = \theta^{\frac{1}{a}}$ for some $a > 0$. Consider the function ϕ, where

$$\phi(t) = t^m.$$

Consider the sequence

$$A_1, A_2, \ldots, A_n \ldots,$$

where

$$A_n = \frac{1}{n} \cdot \sum_{i=1}^{n} \phi(E_{F^n_{y_i}}\ \theta),$$

and y_i identifies the ith interval in

$$[x_0^n, x_1^n], (x_1^n, x_2^n], \ldots, (x_i^n, x_{i+1}^n], \ldots, (x_{n-2}^n, x_{n-1}^n], (x_{n-1}^n, x_n^n],$$

and

$$x_{i+1}^n - x_i^n = \frac{1}{n} \text{ for all } i \in \{0, 1, 2, \ldots, n-1\}.$$

Then for all $a > 0$:

[**1**] the sequence is strictly increasing (i.e., $n' > n \geq 1$ implies $A_{n'} > A_n$) if $m > 1$ or $m < 0$;

[**2**] the sequence is strictly decreasing (i.e., $n' > n \geq 1$ implies $A_{n'} < A_n$) if $0 < m < 1$;

[**3**] the sequence is constant if $m = 1$ or $m = 0$.

3.2 Two questions suggested by Theorem N1

3.2.1 Is Theorem N1 implied by basic propositions concerning stochastic orders?

One might expect the general theory of stochastic orders to contain propositions that imply Theorem N1. The function $\phi(t) = t^m$ of Theorem N1 is convex for $m > 1$ and $m < 0$ and concave for $0 < m < 1$. If $m = 1$ the function is both (weakly) concave and (weakly) convex. That suggests that propositions about the *convex order* might be particularly fruitful.

The convex order is studied in Chapter 3 of Shaked and Shanthikumar, *Stochastic Orders* (2007). Given two random variables X and Y, one says that X is smaller than Y in the convex order, if for any convex function $\phi : I\!R \to I\!R$:

$$E[\phi(X)] \leq E[\phi(Y)]$$

whenever the two expectations exist. Consider the case $m = 1, \phi(t) = t$. Let X be a random variable taking n values, namely the mean of each of the equal-probability intervals in the n-interval structure. Let Y be a random variable taking $n+1$ values, namely the mean of each of the equal-probability intervals in the $(n+1)$-interval structure. We could immediately conclude (using the terminology of Theorem N1) that $A_n \leq A_{n+1}$ if we could establish that X is smaller than Y in the convex order. It is shown in Shaked and Shantikumar (Theorem 3.A.2) that this is the case if and only if

$$E[|X - a|] \leq E[|Y - a|] \text{ for all } a \in I\!R.$$

Now let $a = \frac{1}{2}$ and $n = 2$. Let X equal $1/4$ or $3/4$, each with probability $1/2$, and let Y equal $1/6, 1/6$ and $5/6$ each with probability $1/3$. Then $E[|X - a|]$ is $\frac{1}{2} \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{4}$, but $E[|Y - a|]$ is $\frac{1}{3} \cdot \frac{1}{3} + \frac{1}{3} \cdot 0 + \frac{1}{3} \cdot \frac{1}{3} = \frac{2}{9} < \frac{1}{4}$.

A search among other stochastic-order propositions has so far discovered none that have direct application to Theorem N1.

3.2.2 Do the conclusions of Theorem N1 hold for a much wider class of distributions on $[0, 1]$?

One might hope that the specialized intricacies in the proof of Theorem N1 can be avoided. Could there be a more general proof that holds for a wider class of distributions on $[0, 1]$? A first step in answering this is to find a counterexample in which the distrib'ution does not belong to the class considered in the Theorem and the conclusions of the Theorem do not hold. We now construct such a counterexample

Consider the payoff function

$$u = q\theta - \frac{2}{3} \cdot q^{3/2}.$$

Given a signal y, the best quantity is

$$[E(\theta \mid y)]^2$$

and maximal expected payoff is

$$\frac{1}{3} \cdot [E(\theta \mid y)]^3.$$

Compare an information structure which divides $\Theta = [0, 1]$ into two equal-probability intervals with a structure which divides Θ into three equal-probability intervals. Theorem N1 tells us that if the prior has the form $G(\theta) = \theta^{\frac{1}{a}}$ for $a > 0$, then for the payoff function just introduced, the average best payoff over the three intervals of the second structure is higher than the average best payoff over the two intervals of the first structure. It also tells us that we have Complements, i.e., the average best quantity for the second structure is higher than the average best quantity for the first structure.

We shall present another distribution on $[0, 1]$ for which (1) it remains true (for the payoff function just introduced) that average best payoff rises when we go from two equal-probability intervals to three, but (2) we now have Substitutes, i.e., average best quantity goes down. The example is "reasonable" in the sense that the Producer benefits when we go from two intervals to three. The details are given in the Appendix B.

3.3 A non-Blackwell Information Gatherer who performs a scale/location transform on a base distribution.

In another paper (Marschak, Shanthikumar, Zhou, November 2013) we consider an IG who has a "base" distribution on the states. After exerting effort, he sends a signal to the Producer. The signal identifies a scale/location transform of the base distribution. When IG effort increases, there is a drop in the average scale parameter, over all the signals. The Producer has a prior distribution on the states. That prior is not the same as the IG's base distribution. The prior has the property that once the Producer receives the IG's signal, his posterior is precisely the transform of the base distribution which the signal identifies.

For an example, let the state set be $[0,1]$ and let the Producer's prior be uniform on $[0,1]$. Let the IG's base distribution be uniform on $[-\frac{1}{2}, \frac{1}{2}]$. The signal y identifies the location coefficient $a(y)$ and the scale coefficient $b(y)$. That pair yields a transform of the base distribution, namely the uniform distribution on $\left[a(y) - \dfrac{b(y)}{2}, a(y) + \dfrac{b(y)}{2}\right] \subset [0,1]$, and that becomes the Producer's posterior. The IG has a finite signal set Y. To complete our description of the IG we have to specify the marginal probability of each signal y. We let that marginal probability equal the width of $\left[a(y) - \frac{b(y)}{2}, a(y) + \frac{b(y)}{2}\right]$ divided by the width of $[0,1]$. That equals $b(y)$. So the average of the scale parameters is $\sum_{y \in Y} \ [b(y)]^2$. If that average drops when the IG switches from one structure to another, then the new structure requires more effort.

This IG is not a Blackwell IG, because there is no ranking of structures such that a higher-ranking structure is at least as useful as a lower-ranking one for every Producer, no matter what the prior and the payoff function may be.

But this IG's structures do articulate well with a certain type of Producer, namely the classic price-taking "newsvendor". The state θ is the unknown demand for a product whose price is known to be one. The newsvendor has to place an order q before demand is known. The cost is c per unit, where $0 < c < 1$. Thus the Producer has the payoff function $u(q,\theta) = \min(q,\theta) - cq$ and responds to a signal y by choosing the order $\hat{q}(y)$ which maximizes $E(\min(q,\theta)|y) - cq$. That maximizer is not a function of $E(\theta|y)$ and is neither convex nor concave in the posterior distribution.

We are interested in the effect of higher IG effort on the Producer's highest attainable expected payoff; on the expected value, over all signals, of $\hat{q}(y)$; and on the expected value, over all signals, of the *ex post* quantity delivered to buyers, i.e. $E_{F_y} \min(\hat{q}(y), \theta)$. We show the following: (1) higher IG effort increases the average of the highest attainable expected value of u, averaged over all the IG's signals; (2) there is a critical value of c such that on one side the average order rises when effort rises and on the other side the average order drops; (3) regardless of c, the average *ex post* quantity rises when effort rises.[19]

We again find — just as we did for the IG/Producer pairs studied in the preceding non-Blackwell section — that the techniques needed to obtain our results are specialized to our particular IG and our particular Producer. The generality of the Blackwell Theorems is not available.

[19]These results generalize. The classic newsvendor suffers a penalty when the demand turns out not to equal the order he placed. The penalty is linear and symmetric in the discrepancy. Our results generalize to a newsvendor for whom the penalty is nonlinear and asymmetric .

4 Some Complements/Substitutes results that are implied by the preceding Theorems.

4.1 A Blackwell IG and a Producer whose best quantity depends on the posterior mean: applying the Blackwell Theorems and Theorems B1-B3.

Now consider a Producer whose payoff function has the form

$$u(q, \theta) = \theta \cdot L(q) + M(q),$$

where $q \in I\!R^+$ and the state θ lies in the state set $\Theta \subseteq I\!R$. For a fixed q, the expected value of u depends only on the expected value of θ; additional information about the distribution of θ is not needed. If this Producer receives a signal y from the Information-gatherer, then he chooses a quantity which depends only on the posterior mean $E_{F_y}\ \theta$.

Applying the Blackwell theorems Suppose we have found a function $q^* : I\!R \to I\!R^+$ such that for every posterior mean $E_{F_y}\ \theta$, the Producer's unique best quantity is $q^*(E_{F_y}\ \theta)$. Suppose, moreover, that the function q^* is convex. Then q^* is also a convex function on the set of possible distributions on Θ, since a distribution and its mean are linearly related. Let the Information-gatherer be a Blackwell IG and let us turn to the Blackwell theorems. Consider the statement "(a) implies (c)" which appears in all of them, where (a) says that the experiment $\mathcal{E}'$ is at least as informative as the experiment $\mathcal{E}$, while (c) says that for any convex function on the IG's posteriors, the average value of that function (over all signals) is at least as high for $\mathcal{E}'$ as for $\mathcal{E}$.

We can conclude, in view of the Fourth Blackwell Theorem, that for this Producer, and for any Blackwell IG, we have Complements: if the IG works harder, switching from the experiment $\mathcal{E}$ to the at-least-as-informative experiment $\mathcal{E}'$, then the average quantity chosen by the Producer cannot fall. If, on the other hand, we find the function q^* to be concave, then we have Substitutes: when the IG works harder the average quantity chosen by the Producer cannot rise.

How can we establish the convexity or concavity of the function q^*? We can use the sufficient conditions provided in Theorems B6 and B7. Instead we shall exploit the special properties of the payoff function $u = \theta L(q) + M(q)$ and we will apply standard comparative-statics techniques.

We will find that when the Producer is a price-taker who is uncertain about price, or a monopolist who is uncertain about a linear demand curve, then if his cost function is thrice differentiable the sign of its third derivative plays a key role in determining whether we have Complements or Substitutes. If the Producer is uncertain about cost, then the convexity of the reciprocal of marginal cost plays a key role.

Let us use the abbreviation w for the posterior mean and assume that L and M are differentiable. The quantity $q^*(w)$ is the unique maximizer of

$$\pi(q,w) = wL(q) + M(q).$$

It is the unique solution to the first-order condition

$$wL'(q) + M'(q) = 0.$$

We now show that:

(**A1**) q^* is (weakly) increasing if $L'(q) > 0$ for all $q \geq 0$.

(**A2**) q^* is (weakly) decreasing if $L' < 0$ for all $q \geq 0$.

To establish these statements, note that $\frac{\partial^2 \pi(q,w)}{\partial q \partial w} = L'(q)$. If $L'(q) > 0$ for all $q \geq 0$, then π is supermodular and a standard result[20] tells us that q^* is (weakly) increasing. Suppose, on the other hand that $L'(q) < 0$ for all $q \geq 0$. Define a new variable $\tau = -w$ and consider $\frac{\partial^2 \pi(q,-\tau)}{\partial q \partial \tau} = -L'(q) > 0$. Then we again have supermodularity. The function $q^*(-\tau)$ is (weakly) increasing in τ and hence $q^*(w)$ is (weakly) decreasing in w.

Having established (**A1**),(**A2**), we shall now use the following two facts:

(**B**1) For an increasing function f, the inverse is convex if f is concave and the inverse is concave if f is convex.

(**B**2) For a decreasing function g, the inverse is convex if g is convex and the inverse is concave if g is concave.

Now let h denote the inverse of q^*. Since $q^*(w)$ is the unique solution to the first-order condition $L'(q) \cdot w + M'(q) = 0$, we have

$$w = h(q^*(w)),$$

where

$$h(q) = -\frac{M'(q)}{L'(q)} \quad \text{provided } L'(q) \neq 0.$$

Applying (**A**1), (**A**2),(**B**1), (**B**2), we obtain the following.

<u>**C**. Suppose $L'(q) > 0$ for all $q \geq 0$.</u>

Then

[20]See Milgrom and Shannon, 1994.

($\mathbf{C1}$) The function q^* and its inverse $h = -\frac{M'}{L'}$ are increasing.

($\mathbf{C2}$) q^* is convex if h is concave and q^* is concave if h is convex.

D. Suppose $L'(q) < 0$ for all $q \geq 0$.

Then

($\mathbf{D1}$) q^* and its inverse $h = -\frac{M'}{L'}$ are decreasing.

($\mathbf{D2}$) q^* is convex if h is convex and q^* is concave if h is concave.

We now apply ($\mathbf{B}$1), ($\mathbf{B}$2), ($\mathbf{C}$) and ($\mathbf{D}$) to several types of Producers.

4.1.1 A price-taking Producer who is uncertain about price.

This Producer's payoff function is

$$u(q, \theta) = \theta q - C(q),$$

where price is $\theta \in \Theta \subseteq I\!R^+$ and any nonnegative quantity q can be chosen. So in the general framework just discussed, the term $L(q)$ is q and $M(q)$ is $-C(q)$. Assume that C is twice differentiable, with $C(0) = 0$ and $C'(q) > 0, C''(q) > 0$ for all $q \geq 0$. Then given the posterior mean w, the Producer chooses $q^*(w)$, the unique maximizer of $\pi(q, w) = L(q) \cdot w + M(q)$. Since $L'(q) = 1 > 0$ for $q \geq 0$ we can apply ($\mathbf{C}$1), ($\mathbf{C}$1). Note that

$$h(q) = -\frac{M'(q)}{L'(q)} = C'(q).$$

Suppose now that the function C is thrice differentiable. Then, using ($\mathbf{C}$1), ($\mathbf{C}$2), we conclude that the function $q^* : I\!R^+ \to I\!R^+$ is concave if $C'''(q) > 0$ for all $q \geq 0$ and convex if $C'''(q) < 0$ for all $q \geq 0$. Hence for this Producer and any Blackwell IG whose structures satisfy the assumptions of the Fourth Blackwell theorem:

- we have Substitutes if $C'''(q) > 0$ for all $q \geq 0$.
- we have Complements if $C'''(q) < 0$ for all $q \geq 0$.

In particular if

$$C(q) = r \cdot \frac{1}{1+k} q^{1+k}, r > 0, k > 0,$$

then we have Substitutes if $k > 1$ and we have Complements for $0 < k < 1$.

4.1.2 A monopolist who is uncertain about demand.

The Producer is now a monopolist whose payoff function is

$$u(q,\theta) = qD(q,\theta) - C(q).$$

The state variable θ is nonnegative. The cost function C is twice differentiable and satisfies $C(0) = 0, C' > 0, C'' > 0$ for all $q \geq 0$. Let the demand function D be linear. Then, just as in the price-taker case that we just studied, there are functions L, M such that the expected payoff for the posterior mean w is

$$\pi(q,w) = wL(Q) + M(q).$$

First scenario: the demand curve rotates, with the quantity intercept fixed.

Suppose

$$\text{price } = D(q,\theta) = 1 - \theta q.$$

Given a posterior mean w, the Producer maximizes

$$\pi(q,w) = \left(-q^2\right) \cdot w + (q - C(q)).$$

So $L(q) = -q^2$ and $M(q) = q - C(q)$. For every $w > 0$, there is a unique maximizer $q^*(w)$. Since $L' < 0$, (**D**1) tells us that q^* is decreasing. Make the further assumption that $C'(0) = 0$. Then the first-order condition for a maximizer of $\pi(\ \cdot\ , w)$ tells us that $1 - C'(q^*(w)) > 0$. For the inverse function h we have

$$h(t) = -\frac{M'(t)}{L'(t)} = \frac{1 - C'(t)}{2t}$$

and

$$h''(t) = \frac{-C'''(t)t^2 + 2tC''(t) + 2(1 - C'(t))}{2t^3}.$$

Since $1 - C'(q^*(w)) > 0, C'' > 0$, we conclude that if $C''' < 0$, then the inverse of the function q^* is convex and hence, by (**D**2), q^* is also convex.

So for this Producer and any Blackwell IG whose structures satisfy the assumptions of the Fourth Blackwell Theorem:

If $C'''(q) < 0$ for all $q > 0$, then we have Complements.

For an example, consider the cost function $C(q) = r \cdot \frac{q^{1+k}}{1+k}, k \geq 0, r > 0$. Then $C'''(q) = rk(k-1)q^{k-2}$, which is negative when $0 < k < 1$. So

if $k \in (0,1)$, we have Complements.

Second scenario: the demand curve rotates, with the price intercept fixed.

Now we have

$$\text{price } = D(q, \theta) = \theta(1 - q).$$

Given the posterior mean w, the Producer maximizes

$$\pi(q, w) = w \cdot (1 - q) \cdot q - C(q).$$

Assume that C is thrice differentiable and, as before, that $C' > 0, C'' > 0$. Make the additional assumption that $C'(0) = 0$. We permit the Producer to choose any nonnegative q. Note that for every $w \geq 0$ there is a unique maximizer of $\pi(q, w)$. As before, the maximizer is denoted $q^*(w)$. The first-order condition is

$$w \cdot (1 - 2q^*(w)) - C'(q^*(w)) = 0$$

Note that $C' > 0$, so $1 - 2q^*(w) > 0$ and for any $w \geq 0$, the maximizer $q^*(w)$ lies in $[0, \frac{1}{2})$. Then $h(t) = \frac{C'(t)}{1-2t}, t \in [0, \frac{1}{2})$ is the inverse of the function q^*. We have

$$h'(t) = \frac{(1 - 2t)C'' + 2C'}{(1 - 2t)^2} > 0,$$

since $1 - 2t > 0, C' > 0, C'' > 0$. We then obtain

$$h''(t) = \frac{(1 - 2t)^2 C''' + 4(1 - 2t)C'' + 8C'}{(1 - 2t)^3}.$$

If $C''' > 0$, then $h'' > 0$, so h is convex on $[0, \frac{1}{2})$. Therefore (by (**B**1)), q^* is increasing and concave if $C''' > 0$. We conclude that for any Blackwell IG whose structures satisfy the assumptions of the Fourth Blackwell Theorem:

we have Substitutes if $C'''(q) > 0$ for all $q \geq 0$.

In particular if $C(q) = r \cdot \frac{1}{1+k} q^{1+k}, r > 0, k \geq 0$, then

we have Substitutes if $k > 1$.

Third scenario: the demand curve maintains its slope but it shifts up and down.

Finally, suppose that

$$\text{price } = D(q, \theta) = \theta - q.$$

Then

$$\pi(q, w) = qw - q^2 - C(q).$$

So $L(q) = q, M(q) = -q^2 - C(q)$. We have $L'(q) = 1 > 0$, $h(q) = -\frac{M'q)}{L'(q)} = 2q + C'(q)$, and

$$h''(q) = C'''(q).$$

Applying (**C**1), we see that the unique maximizer $q^*(w)$ is increasing in w. Since h is convex if $C''' > 0$ and concave if $C''' < 0$, we again obtain (using (**C**2)) the same result as we obtained for the price-taker: for this Producer and for any Blackwell IG whose structures satisfy the assumptions of the Fourth Blackwell Theorem:

- we have Substitutes if $C'''(q) > 0$ for all $q \geq 0$.
- we have Complements if $C'''(q) < 0$ for all $q \geq 0$.

In particular if $C(q) = r \cdot \frac{q^{1+k}}{1+k}, r > 0, k \geq 0$,then $C''' = rk \cdot (k-1)q^{k-2}$ and

we have Substitutes if $k > 1$ and Complements if $k < 1$.

4.1.3 price-taker who is uncertain about cost.

Now consider a Producer who sells at a fixed price (which we take to be one) and has total cost function $\theta C(q)$. As before, assume $C' > 0, C'' > 0$. Again assume that $C'(0) = 0$. Given a posterior mean w, the producer maximizes

$$\pi(q, w) = (-C(q)) \cdot w + q.$$

So $L(q) = -C(q), M(q) = q$, and $h(q) = -\frac{M'}{L'} = \frac{1}{C'(q)}$. For every $w > 0$, we see that $q^*(w)$, the unique maximizer of π is positive. Since $L' < 0$ for all $q > 0$, we apply (**C**2). We conclude that for this Producer and any Blackwell IG whose structures satisfy the assumptions of the Fourth Blackwell Theorem

we have Complements if $\frac{1}{C'(q)}$ is convex.

In particular if $C(q) = r \cdot \frac{q^{1+k}}{1+k}, r > 0, k > 0$, then $\frac{1}{C'(q)} = \frac{1}{rq^k}$ is convex for $q > 0$. So:

we have Complements for all $k > 0$.

Note that there is no function $C : I\!R^+ \to I\!R^+$ with $C' > 0$ such that $\frac{1}{C'}$ is concave.[21] So if this Producer uses a Blackwell IG, and every positive number is a possible posterior mean, then we cannot have Substitutes.

[21]Suppose $\frac{1}{C'}$ were a concave function, say f, on $I\!R^+$. Then f is bounded from above by any tangent line. The tangent line is negatively sloped and the absolute value of its slope becomes arbitrarily large as t increases without limit. Hence we have $\lim_{t\to\infty} f(t) = -\infty$. That contradicts our assumption that $C' > 0$.

4.1.4 Applying Theorems B1-B4 or Theorems $\tilde{B}1 - \tilde{B}4$ when the IG's structures are linear combinations of two anchor structures.

If the IG has a family of structures I_k indexed by $k \in [0,1]$, and I_k is a linear mixture of an anchor structure with the null property and another anchor structure, then all the preceding Complements/Substitutes results can be obtained without using the Blackwell theorems even though this IG is a Blackwell IG. Instead, we first establish, for each result, that the Producer's best quantity is a convex or concave function of the posterior mean. We then apply Theorems B1-B4 if the state and signal sets are finite, and Theorems $\tilde{B}1 - \tilde{B}4$ if they are not.

4.2 Applying Theorems N1-N3: a non-Blackwell IG partitions the states into equal-probability or equal-length intervals and conveys signals to a Producer whose best quantity depends on the posterior mean.

The IG in Theorems N1-N3 has n signals $y_1, \ldots, y_n$. Signal y_i tells the Producer that the state lies in the subinterval of the state set $[0,1]$ identified by that signal and that the posterior is $F^n_{y_i}$ with mean $E_{F^n_{y_i}}\ \theta$. Higher n means more IG effort. Each theorem makes an assumption about the prior and considers the effect of an increase in n on the average, over all n signals, of $\left[E_{F^n_{y_i}}\ \theta\right]^m$. Note that if that average rises (drops), then so does the average of $r \cdot \left[E_{F^n_{y_i}}\ \theta\right]^m$, where $r > 0$.

Now consider a price-taking Producer whose total cost for the quantity q is

$$\frac{1}{1+k} \cdot q^{1+k},$$

where $k > 0$. Price is $\theta \in [0,1]$. Given the signal y, the Producer's best (expected-payoff-maximizing) quantity is $(E(\theta|y))^{1/k}$. So the "m" which appears in the three theorems is $\frac{1}{k} > 0$. The three theorems tell us that

we have Complements if $k < 1$ and Substitutes if $k > 1$.

Now consider the value of the IG's effort to the Producer, i.e., the average, over all n signals y, of

$$E(\theta|y) \cdot \left(E(\theta|y)^{1/k}\right) - \frac{1}{1+k}\left(E(\theta|k)^{1/k}\right)^{1+k} = \frac{k}{1+k} \cdot [E(\theta|y)]^{\frac{1}{k}+1}.$$

Applying the theorems to the case $m = \frac{1}{k} + 1 > 1$, we see that the Prodcuer strictly benefits when the IG works harder.

Now suppose that price is one but cost is

$$\theta \cdot \frac{1}{1+k} \cdot q^{1+k},$$

where $k > 0$. Given the signal y, the Producer's best quantity is $(E(\theta|y))^{-1/k}$. So the "m" which appears in the three theorems is now $\frac{-1}{k} < 0$. The three theorems tell us that

we have Complements.

The value of the IG's effort to the Producer is now the average, over all n signals y, of

$$\frac{k}{k+1} \cdot [E(\theta|y)]^{-1/k}.$$

Applying the theorems to the case $m = -\frac{1}{k} < 0$, we again see that the Producer strictly benefits when the IG works harder.

Notice that whenever the equal-probability or equal-length partitioner moves from a t-interval partitioning to an ℓt-interval partitioning, where ℓ is a positive integer, the move is a refinement and so it can never damage the Producer. That means we can apply the Fourth Blackwell Theorem to the move. If the Producer's best quantity is a convex (concave) function of the posterior, then we have Complements, i.e., the move leads to a not-lower average best quantity (we have Substitutes, i.e., the move leads to a not-higher average best quantity). In the preceding results the best quantity, or the value, are either convex or concave in the posterior. So for that sort of a move by the IG, we can derive our results using the Fourth Blackwell Theorem. We cannot do so, of course, when the IG moves from a partitioning with n intervals to one with $n' > n$ intervals and $\frac{n'}{n}$ is not an integer. In that case the convexity or concavity of best quantity as a function of the posteriors is of no help.

4.3 An application of the second strict-garbling theorem.

Consider again the price-taking Producer whose payoff function is

$$\theta q - \frac{1}{1+k} q^{1+k},$$

where $k > 0$. Given a signal y, the Producer's best quantity is $[E(\theta|y)]^{1/k}$, which is *strictly* convex in the posterior mean.

Suppose that the state set is $\{\theta_1, \ldots, \theta_n\} \in I\!R^+$, where the n states are distinct, and that all n probabilities are positive. Next consider the n-signal "one-point forecaster" studied above in Remark 2.3.5.3. Suppose he moves from effort x to effort x', where $\frac{1}{n} \leq x < x' \leq 1$. We claim that the Second Strict Garbling Theorem tells us that we have *strict* Complements: the average best quantity for x' is strictly larger than the average best quantity for x. Condition (C1) of the theorem is met, since we explicitly exhibit the garbling matrix. All entries in that matrix are positive, which means that $(*)$ of (C2) is satisfied for every j, u, v. If $(**)$ of (C2) were violated for every j, u, v, then all n posterior means would be equal. One can show that this means all n states would be identical, contrary to assumption. That concludes the argument.

5 Concluding remarks.

Our most striking Complements/Substitutes results arise when the Producer's best quantity depends on the posterior mean. We examined a number of cases in which (1) the state θ enters the payoff function in a simple way as a multiplicative parameter, and (2) the best quantity is a convex (concave) function of the posterior mean (and hence of the posterior itself). We then get sharp Complements/Substitutes results for any Blackwell IG. The critical role of the sign of the third derivative of the cost function is an unexpected result, not suggested by any simple intuition.

It remains challenging to obtain results for Blackwell IG's when the Producer's best quantity is *not* a convex (concave) function of the posterior. For such Producers, statement (c) in the Blackwell theorems is useless and a fresh analysis of the Producer's response to a Blackwell IG's signals is needed. We have not studied such situations in this paper.

We have, however, studied non-Blackwell IGs. Special techniques are needed for the non-Blackwell IG who partitions the state set into n equal-probability or equal-length intervals and can choose any positive integer n. Special techniques are also needed for the non-Blackwell IG who performs a scale/location transform of a base distribution. For the equal-probability partitioner and certain Producer payoff functions we get Complements/Substitutes results that duplicate what we found when the IG is a Blackwell IG.

A large and varied terrain remains to be explored. We may, for example, let our Producer be risk-averse, so that he maximizes the expected value of a concave function of what we have called payoff. In particular, the Producer might respond to the IG's signal by maximizing a linear combination of the posterior mean of payoff and the posterior variance of payoff. Might that reverse some of our Complements/Substitutes results?

If the Producer chooses the IG's effort, pays for it, and knows that he gets the effort he pays for, then the Complements/Substitutes question concerns the effect of lower effort prices on the average quantity produced. Alternative price schedules have to be studied. We made a start for the case of our "k-mixture" Blackwell IG, where the price charged for the effort k has to rise sufficiently rapidly if the Producer's best k is going to be in the interior of $[0, 1]$. We would expect interior choices to be made in real settings, so it is important to study effort pricing for other IGs, where structures are again indexed and each value of the index again carries a price. If the Producer cannot be sure what IG effort his payment buys, then we enter the difficult realm of incentive models. Introducing incentives in the IG/Producer model is a distant challenge, since so much still remains to be learned about the incentive-free case.

We conclude by emphasizing again the strong empirical motivation for building a Complements/Substitutes theory. To explain and predict the occupational shifts and the productivity effects of the ongoing information-technology revolution requires some clarifying theory. One path is to study complex models in which one traces the effect of improved information-gathering on the structure of many-person organizations and the composition of their workforce. A simpler path is our two-person model. Under specific assumptions about the IG and the Producer, we predict the direction in which production quantities move when informational effort increases. Empiricists may be able to check the assumptions and to test the predictions.

APPENDIX

A Proofs

A.1 Proof of the first strict-garbling theorem in section 2.2.

Let $Q = (\vec{q_1}, \ldots, \vec{q_m})$, $Q' = (\vec{q'_1}, \ldots, \vec{q'_{m'}})$ denote, respectively, the n-by-m matrix of posterior state probabilities for the experiment $\mathcal{E}$ and the n-by-m' matrix of posterior state probabilities for the experiment $\mathcal{E}'$. We start by establishing the following two facts about Q, Q':

(F1) There exists an m'-by-m column-stochastic matrix $\bar{B} = ((\bar{b}_{ts}))$ such that

$$(1) \quad Q = Q' \cdot \bar{B},$$

i.e.,

$$(2) \text{ for every } s \text{ in } \{1, \ldots, m\} \text{ we have } \vec{q_s} = \sum_{t=1}^{m'} \vec{q'_t} \cdot \bar{b}_{ts}.$$

(The posterior-probability vector $\vec{q_s}$ in the experiment E is a weighted average of all the experiment-E' posterior-probability vectors, with weights equal to the sth column of $\bar{B}$).

(F2) For every s in $\{1, \ldots, m\}$ and every t in $\{1, \ldots, m'\}$ we have

$$\bar{b}_{ts} = b_{ts} \cdot \frac{\Pr(y'_t)}{\Pr(y_s)}.$$

To establish **(F1)**,**(F2)**[22] , we shall use the following three diagonal matrices: D^p which has the state probabilities $p_1, \ldots, p_n$ on the diagonal; D, which has the experiment-E signal probabilities $\Pr(y_1), \ldots, \Pr(y_m)$ on the diagonal; and D', which has the experiment-E' signal probabilities $\Pr(y'_1), \ldots, \Pr(y'_{m'})$ on the diagonal. In view of the above equalities

$$q_{is} = \frac{p_i \cdot \lambda_{is}}{\sum_{i=1}^{n} p_i \cdot \lambda_{is}}; \ q'_{it} = \frac{p_i \cdot \lambda'_{it}}{\sum_{i=1}^{n} p_i \cdot \lambda'_{it}},$$

[22] **(F1)** is the same as Theorem 8.2 in Marschak/Miyasawa, while **(F2)** is shown in the proof of that theorem (equation (8.23)). The proof uses a number of results that precede it. We prefer a self-contained argument.

we may write s

$$D^p \cdot \Lambda = Q \cdot D; D^p \cdot \Lambda' = Q' \cdot D'$$

or, equivalently,

$$\Lambda = (D^p)^{-1} \cdot Q \cdot D; \Lambda' = (D^p)^{-1} \cdot Q' \cdot D'.$$

Since (by assumption) $\Lambda = \Lambda' \cdot B$, we have

$$(D^p)^{-1} \cdot Q \cdot D = (D^p)^{-1} \cdot Q' \cdot D' \cdot B.$$

Pre-multiplying both sides by D^p and then post-multiplying by D^{-1}, we obtain

$$Q = Q' \cdot \bar{B}, \text{ where } \bar{B} = D' \cdot B \cdot D^{-1}.$$

But $\bar{B} = D' \cdot B \cdot D^{-1}$ is equivalent to **(F2)**. Moreover, since $Q = Q' \cdot \bar{B}$ is equivalent to condition (2) in **(F1)**, and since every vector $\vec{q}_s, \vec{q'_t}$ sums to one, we see that $\bar{B}$ is indeed column-stochastic. So **(F1), (F2)** are established.

Note also that since $D^p \cdot \Lambda' = Q' \cdot D$, Condition (C3) implies[23]

$$\text{rank of } \Lambda' = \text{rank of } Q = m' \tag{3}$$

In view of (2) in **(F1)**, we may now write

$$V_\phi(\Lambda) = \sum_{s=1}^{m} \ (\Pr(y_s)) \cdot \phi \left(\sum_{t=1}^{m'} \ \vec{q'_t} \cdot \bar{b}_{ts} \right). \tag{4}$$

First suppose that ϕ is strictly convex. Then the following hold:

$$\text{for any } s \text{ in } \{1, \ldots, m\} \text{ we have } \phi \left(\sum_{t=1}^{m'} \ \bar{b}_{ts} \cdot \vec{q'_t} \right) \leq \sum_{t=1}^{m'} \ \bar{b}_{ts} \cdot \phi(\vec{q'_t}). \tag{5a}$$

$$\text{for at least one } s \text{ in } \{1, \ldots, m\} \text{ we have } \phi \left(\sum_{t=1}^{m'} \ \bar{b}_{ts} \cdot \vec{q'_t} \right) < \sum_{t=1}^{m'} \ \bar{b}_{ts} \cdot \phi(\vec{q'_t}). \tag{5b}$$

The weak inequality in (5a) follows from Jensen's inequality directly, while to show strict inequality in (5b), we use the strict Jensen's inequality which holds for a strictly convex function.[24]Consider the integers j, u, v which appear in condition (C2). The inequality in (5b) holds for $s = j$ because:

---

[23]This is the same as Theorem 9.2 in Marschak/Miyasawa.

[24]In general, let f be a strictly convex function from C to $I\!R$, where C is a convex set in $I\!R^k$. Then for any $x_1, \ldots, x_N \in C$, and any $(\lambda_1, \ldots, \lambda_N)$ with $\sum_{1=1}^{N} \ \lambda_i = 1$ and $\lambda_i \geq 0$ for all $i \in \{1, \ldots, N\}$, we have

$$f(\lambda_1 \cdot x_1 + \cdots + \lambda_N \cdot x_N) < \lambda_1 \cdot f(x_1) + \cdots + \lambda_N \cdot f(x_N)$$

provided that: there exist i, j, with $i \neq j$, such that $\lambda_i > 0, \lambda_j > 0$ and $x_i \neq x_j$.

(i) ϕ is strictly convex.

(ii) Since $\bar{B}$ is column-stochastic, we have $\bar{b}_{1j} + \cdots + \bar{b}_{mj} = 1$.

(iii) **(F2)** and (C2) tell us that $u \neq v, \bar{b}_{uj} > 0, \bar{b}_{vj} > 0$.

(iv) By (3), any two columns of Q' are distinct, in particular, the columns $\vec{q'_u}$ (which has weight $\bar{b}_{uj} > 0$ in the inequaliy in (5)), and $\vec{q'_v}$ (which has weight $\bar{b}_{vj} > 0$).

Now (4). (5a) and (5b) imply

$$V_\phi(\Lambda) < \sum_{s=1}^{m} \Pr(y_s) \cdot \left[\sum_{t=1}^{m'} \bar{b}_{ts} \cdot \phi(\vec{q'_t}) \right]. \tag{6}$$

Using **(F2)** and (6), we then have

$$\begin{aligned} V_\phi(\Lambda) &< \sum_{s=1}^{m} \Pr(y_s) \cdot \left[\sum_{t=1}^{m'} b_{ts} \cdot \frac{\Pr(y'_t)}{\Pr(y_s)} \cdot \phi(\vec{q'_t}) \right] = \sum_{s=1}^{m} \left[\sum_{t=1}^{m'} b_{ts} \Pr(y'_t) \cdot \phi(\vec{q'_t}) \right] \\ &= \left(\sum_{t=1}^{m'} \Pr(y'_t) \cdot \phi(\vec{y'_t}) \right) \cdot \sum_{s=1}^{m} b_{ts} = \sum_{t=1}^{m'} \Pr(y'_t) \cdot \phi(\vec{q'_t}) = V_\phi(\Lambda'). \end{aligned}$$

(The next-to-last equality uses the fact that B is row-stochastic).

That completes the proof for the case where ϕ is strictly convex. There is an analogous proof for the strictly concave case. □

A.2 Proof of the second strict-garbling theorem in section 2.2.

The proof repeats all steps of the preceding proof up to (and including) the sentence "So **(F1)**,**(F2)** are established". After that, the proof proceeds in the following way.

Note that (2) of **(F1)** tells us that for every $s \in \{1, \ldots, m\}$ we have

$$E_{\vec{q_s}}\ \theta = \sum_{t=1}^{m'} \bar{b}_{ts} \cdot \left(E_{\vec{q'_t}}\ \theta \right).$$

So we may write

$$V^*_\phi(\Lambda) = \sum_{s=1}^{m} \Pr(y_s) \cdot \phi \left(\sum_{t=1}^{m'} \bar{b}_{ts} \cdot \left(E_{\vec{q'_t}}\ \theta \right) \right). \tag{3}$$

First suppose that ϕ is strictly convex. In view of (3), it will suffice to show that:

$$(4a) \qquad \text{for any } s \text{ in } \{1,\ldots,m\} \text{ we have } \phi\left(\sum_{t=1}^{m'} \bar{b}_{ts} \cdot \left(E_{\vec{q_t'}}\ \theta\right)\right) \leq \sum_{t=1}^{m'} \bar{b}_{ts} \cdot \phi\left(E_{\vec{q_t'}}\ \theta\right).$$

$$(4b) \qquad \text{for at least one } s \text{ in } \{1,\ldots,m\} \text{ we have } \phi\left(\sum_{t=1}^{m'} \bar{b}_{ts} \cdot \left(E_{\vec{q_t'}}\ \theta\right)\right) < \sum_{t=1}^{m'} \bar{b}_{ts} \cdot \phi\left(E_{\vec{q_t'}}\ \theta\right).$$

Again, (4a) follows from Jensen's inequality. Consider the integers j, u, v which appear in condition (C2). The inequality in (4b) holds for $s = j$ because:

(i) ϕ is strictly convex.

(ii) Since $\bar{B}$ is column-stochastic, we have $\bar{b}_{1j} + \cdots + \bar{b}_{mj} = 1$.

(iii) **(F2)** and $(*)$ of (C2) tell us that $u \neq v, \bar{b}_{uj} > 0, \bar{b}_{vj} > 0$.

(iv) Condition $(**)$ of (C2) tells us that $E_{\vec{q_u'}}\ \theta$ (which has weight $\bar{b}_{uj} > 0$ in the inequality in (4)) is not equal to $E_{\vec{q_v'}}\ \theta$ (which has weight $\bar{b}_{vj} > 0$).

Now (3),(4a) and (4b) imply

$$(5) \qquad V_\phi^*(\Lambda) < \sum_{s=1}^{m} \Pr(y_s) \cdot \left[\sum_{t=1}^{m'} \bar{b}_{ts} \cdot \phi\left(E_{\vec{q_t'}}\ \theta\right)\right].$$

Using **(F2)** and (5), we then have

$$\begin{aligned} V_\phi^*(\Lambda) &< \sum_{s=1}^{m} \Pr(y_s) \cdot \left[\sum_{t=1}^{m'} b_{ts} \cdot \frac{\Pr(y_t')}{\Pr(y_s)} \cdot \phi\left(E_{\vec{q_t'}}\ \theta\right)\right] = \sum_{s=1}^{m}\left[\sum_{t=1}^{m'} b_{ts} \Pr(y_t') \cdot \phi\left(E_{\vec{q_t'}}\ \theta\right)\right] \\ &= \left(\sum_{t=1}^{m'} \Pr(y_t') \cdot \phi\left(E_{\vec{y_t'}}\ \theta\right)\right) \cdot \sum_{s=1}^{m} b_{ts} = \sum_{t=1}^{m'} \Pr(y_t') \cdot \phi\left(E_{\vec{q_t'}}\ \theta\right) = V_\phi^*(\Lambda'). \end{aligned}$$

(The next-to-last equality uses the fact that B is row-stochastic).

That completes the proof for the case where ϕ is strictly convex. There is an analogous proof for the strictly concave case. □

A.3 Proof of Theorem B5

We first show that $B_{kk'}$ is row-stochastic, i.e., all its entries are non-negative and

$$B_{kk'} \cdot \mathbf{e}_m = \mathbf{e}_m.$$

Clearly the entries are nonnegative. Using the fact that $\mathbf{q} \cdot \mathbf{e}_m = 1$, we have:

$$\begin{aligned} B_{kk'} \cdot \mathbf{e}_m &= \left[\frac{k}{k'}\mathbf{H}_m + \left(1 - \frac{k}{k'}\right)\mathbf{e}_m \cdot \mathbf{q}\right] \cdot \mathbf{e}_m \\ &= \frac{k}{k'}\mathbf{H}_m \cdot \mathbf{e}_m + \left(1 - \frac{k}{k'}\right) \cdot \mathbf{e}_m \cdot (\mathbf{q} \cdot \mathbf{e}_m) = \frac{k}{k'}\mathbf{e}_m + \left(1 - \frac{k}{k'}\right) \cdot \mathbf{e}_m = \mathbf{e}_m \end{aligned}$$

Next note that since every row of a likelihood matrix sums to one, we have, $\mathbf{e}_n = \Lambda_1 \cdot \mathbf{e}_m = \Lambda_0 \cdot \mathbf{e}_m$. Using that fact, and recalling that $\Lambda_0 = \mathbf{e}_n \cdot \mathbf{q}$, we have:

$$\begin{aligned} \Lambda_{k'} \cdot B_{kk'} &= (k'\Lambda_1 + (1-k') \cdot \Lambda_0) \cdot \left[\frac{k}{k'}\mathbf{H}_m + \left(1 - \frac{k}{k'}\right) \cdot \mathbf{e}_m \cdot \mathbf{q}\right] \\ &= k'\Lambda_1 \cdot \frac{k}{k'}\mathbf{H}_m + (1-k') \cdot \Lambda_0 \cdot \frac{k}{k'}\mathbf{H}_m \\ &\quad + k'\Lambda_1 \cdot \left(1 - \frac{k}{k'}\right) \cdot \mathbf{e}_m \cdot \mathbf{q} + (1-k') \cdot \Lambda_0 \cdot \left(1 - \frac{k}{k'}\right) \cdot \mathbf{e}_m \cdot \mathbf{q} \\ &= k\Lambda_1 + (1-k') \cdot \frac{k}{k'} \cdot \Lambda_0 + k' \cdot \left(1 - \frac{k}{k'}\right) \cdot (\Lambda_1 \cdot \mathbf{e}_m) \cdot \mathbf{q} + (1-k') \cdot \left(1 - \frac{k}{k'}\right) \cdot (\Lambda_0 \cdot \mathbf{e}_m) \cdot \mathbf{q} \\ &= k\Lambda_1 + (1-k') \cdot \frac{k}{k'} \cdot \Lambda_0 + \left[k' \cdot \left(1 - \frac{k}{k'}\right) + (1-k') \cdot \left(1 - \frac{k}{k'}\right)\right] \cdot \mathbf{e}_n \cdot \mathbf{q} \\ &= k\Lambda_1 + \left[(1-k') \cdot \frac{k}{k'} + k' \cdot \left(1 - \frac{k}{k'}\right) + (1-k') \cdot \left(1 - \frac{k}{k'}\right)\right] \cdot \Lambda_0 \\ &= k\Lambda_1 + (1-k) \cdot \Lambda_0 = \Lambda_k \end{aligned}$$

That concludes the proof. □

A.4 Proof of Theorem $\tilde{B}$1.

Recall our notational convention. If a measure μ has a density function m, then $\chi(m)$ is interpreted as $\chi(\mu)$. Then for a measure μ_1 with density function m_1, a measure μ_2 with density function m_2, and $\lambda \in [0,1]$, the symbol

$$\chi\left(\lambda m_1 + (1-\lambda)\cdot m_2\right)$$

is correctly interpreted as

$$\chi\left(\lambda \mu_1 + (1-\lambda)\cdot \mu_2\right).$$

It suffices to show that for any (θ, y)

$$h^k(y)\cdot\chi\left(\frac{g(\theta)\cdot\lambda^k(y|\theta)}{h^k(y)}\right)$$

is convex in k on $[0,1]$.

Consider $k_1, k_2 \in [0,1]$, $\alpha \in [0,1]$, and $k = \alpha k_1 + (1-\alpha)\cdot k_2$. We have to show that

$$(+) \qquad h^k(y)\cdot\chi\left(\frac{g(\theta)\cdot\lambda^k(y|\theta)}{h^k(y)}\right)$$

$$\leq \alpha\cdot h^{k_1}(y)\cdot\chi\left(\frac{g(\theta)\cdot\lambda^{k_1}(y|\theta)}{h^{k_1}(y)}\right) + (1-\alpha)\cdot h^{k_2}(y)\cdot\chi\left(\frac{g(\theta)\cdot\lambda^{k_2}(y|\theta)}{h^{k_2}(y)}\right).$$

By the linearity of λ^k and h^k in k, we have

$$\lambda^k(y|\theta) = \alpha\cdot\lambda^{k_1}(y|\theta) + (1-\alpha)\cdot\lambda^{k_2}(y|\theta)$$

$$h^k(y) = \alpha\cdot h^{k_1}(y) + (1-\alpha)\cdot h^{k_2}(y).$$

We now abbreviate, letting $h^{k_1}, h^{k_2}, \lambda^{k_1}, \lambda^{k_2}$ stand, respectively, for $h^{k_1}(y), h^{k_2}(y), \lambda^{k_1}(y|\theta), \lambda^{k_2}(y|\theta)$. We have:

$$h^k(y)\cdot\chi\left(\frac{g(\theta)\cdot\lambda^k(y|\theta)}{h^k(y)}\right) = \left(\alpha\cdot h^{k_1} + (1-\alpha)\cdot h^{k_2}\right)\cdot\chi\left(\frac{g(\theta)\cdot\left[\alpha\cdot\lambda^{k_1} + (1-\alpha)\cdot\lambda^{k_2}\right]}{\alpha\cdot h^{k_1} + (1-\alpha)\cdot h^{k_2}}\right)$$

$$= \left(\alpha\cdot h^{k_1} + (1-\alpha)\cdot h^{k_2}\right)\cdot\chi\left(\frac{\alpha h^{k_1}}{\alpha\cdot h^{k_1} + (1-\alpha)\cdot h^{k_2}}\cdot\frac{g(\theta)\cdot\lambda^{k_1}}{h^{k_1}} + \frac{(1-\alpha)\cdot h^{k_2}}{\alpha\cdot h^{k_1} + (1-\alpha)\cdot h^{k_2}}\cdot\frac{g(\theta)\cdot\lambda^{k_2}}{h^{k_2}}\right)$$

$$\leq \left(\alpha\cdot h^{k_1} + (1-\alpha)\cdot h^{k_2}\right)\cdot\left[\frac{\alpha h^{k_1}}{\alpha\cdot h^{k_1} + (1-\alpha)\cdot h^{k_2}}\cdot\chi\left(\frac{g(\theta)\cdot\lambda^{k_1}}{h^{k_1}}\right) + \frac{(1-\alpha)\cdot h^{k_2}}{\alpha\cdot h^{k_1} + (1-\alpha)\cdot h^{k_2}}\cdot\chi\left(\frac{g(\theta)\cdot\lambda^{k_2}}{h^{k_2}}\right)\right]$$

$$= \alpha h^{k_1}\cdot\chi\left(\frac{g(\theta)\cdot\lambda^{k_1}}{h^{k_1}}\right) + (1-\alpha)\cdot h^{k_2}\cdot\chi\left(\frac{g(\theta)\cdot\lambda^{k_2}}{h^{k_2}}\right).$$

(The inequality follows from the convexity of χ and the fact that $\frac{\alpha h^{k_1}}{\alpha\cdot h^{k_1} + (1-\alpha)\cdot h^{k_2}}$ and $\frac{(1-\alpha)\cdot h^{k_2}}{\alpha\cdot h^{k_1} + (1-\alpha)\cdot h^{k_2}}$ are nonnegative and sum to one). That completes the proof for the case where χ is convex. There is an analagous proof for the case where χ is concave. □

A.5 Proof of Theorem $\tilde{B}5$.

Here a symbol of the form $\delta_z(A)$, where A is a set, denotes the Dirac measure, i.e.

$$\delta_z(A) = \begin{cases} 1 & \text{if } z \in A \\ 0 & \text{if } z \notin A. \end{cases}$$

We have to exhibit the probability measure p'_y required in the Markov-kernel definition of garbling. But that definition allows two distinct pairs $(Y, \mathcal{Y})$ and $(Y', \mathcal{Y}')$. In the experiment pair we are now considering, we have $(Y, \mathcal{Y}) = (Y', \mathcal{Y}')$. That means that for every $y \in Y$, we have to exhibit a probability measure p_y on $(Y, \mathcal{Y})$ for which the following hold:

(i) For every $A \in \mathcal{Y}$, $p_y(A)$, as a function of y, is measurable with respect to $(Y, \mathcal{Y})$.

(ii) For every $\theta \in \Theta$, every set $A \in \mathcal{Y}$, and every $k \in [0, 1]$, we have

$$\lambda_\theta^k(A) = \int_Y p_y(A) \lambda_\theta^{k'}(dy).$$

Consider the following measure p_y (a convex combination of δ_y and h^0, which are both measures on $(Y, \mathcal{Y})$):

$$p_y(\ \cdot\ ) = \frac{k}{k'} \cdot \delta_y(\ \cdot\ ) + \left(1 - \frac{k}{k'}\right) \cdot h^0(\ \cdot\ ).$$

Note that $p_y(Y) = 1$, since for every $y \in Y$ we have $\delta_y(Y) = 1$ and since $h^0(Y) = 1$ (because h^0 is a probability measure). So p_y is indeed a probability measure on $(Y, \mathcal{Y})$.

Next we show (i). To do so, consider a fixed set $A \in \mathcal{Y}$, and suppose that for some $y \in Y$ we have

$$(+) \qquad p_y(A) = \frac{k}{k'} \cdot \delta_y(A) + \left(1 - \frac{k}{k'}\right) \cdot h^0(A) \equiv J.$$

Observe that the term $\left(1 - \frac{k}{k'}\right) \cdot h^0(A)$ is independent of y. Since $\delta_y(A)$ takes the value one for every $y \in A$ and the value zero for every $y \in Y \setminus A$, either (+) holds for every $y \in A$ (i.e., $\frac{k}{k'} + \left(1 - \frac{k}{k'}\right) \cdot h^0(A) = J$) or else (+) holds for every $y \in Y \setminus A$ (i.e., $\left(1 - \frac{k}{k'}\right) \cdot h^0(A) = J$). Thus the inverse image of J under p_y (with A fixed) is either A or $Y \setminus A$, both of which belong to $\mathcal{Y}$, since $\mathcal{Y}$ is a σ-algebra of subsets of Y. That establishes (i).

Now we turn to (ii). We first show that for all sets $A \in \mathcal{Y}$ and all $\theta \in \Theta$ we have

$$(++) \qquad \int_Y p_y(A) \cdot \lambda_\theta^{k'}(dy) = \frac{k}{k'} \int_Y \delta_y(A) \cdot \lambda_\theta^{k'}(dy) + \left(1 - \frac{k}{k'}\right) \cdot h^0(A).$$

That is the case because

$$\begin{aligned}
\int_Y p_y(A) \cdot d\lambda_\theta^{k'}(y) &= \int_Y \left[\frac{k}{k'} \cdot \delta_y(A) + \left(1 - \frac{k}{k'}\right) \cdot h^0(A)\right] \cdot \lambda_\theta^{k'}(dy) \\
&= \frac{k}{k'} \int_Y \delta_y(A) \cdot \lambda_\theta^{k'}(dy) + \left(1 - \frac{k}{k'}\right) \cdot h^0(A) \cdot \int_Y 1 \cdot \lambda_\theta^{k'}(dy) \\
&= \frac{k}{k'} \int_Y \delta_y(A) \cdot \lambda_\theta^{k'}(dy) + \left(1 - \frac{k}{k'}\right) \cdot h^0(A).
\end{aligned}$$

(For the last equality we use the fact that $\lambda_\theta^{k'}$ is a probability measure on Y).

Next we can write

$$\int_Y \delta_y(A)\lambda_\theta^{k'}(dy) = \int_A \delta_y(A)\lambda_\theta^{k'}(dy) + \int_{Y \setminus A} \delta_y(A)\lambda_\theta^{k'}(dy).$$

Since $\delta_y(A)$ equals one for $y \in A$ and equals zero for $y \in Y \setminus A$, and since

$$\int_A 1 \cdot \lambda_\theta^{k'}(dy) = \lambda_\theta^{k'}(A),$$

we obtain

$$(+++) \qquad \int_Y \delta_y(A)\lambda_\theta^{k'}(dy) = \lambda_\theta^k(A).$$

Using (++) and (+++) we obtain:

$$\begin{aligned}
\int_Y p_y(A)\lambda_\theta^{k'}(dy) &= \frac{k}{k'} \cdot \lambda_\theta^{k'}(A) + \left(1 - \frac{k}{k'}\right) \cdot h^0(A) \\
&= \frac{k}{k'} \cdot \left[k' \cdot \lambda_\theta^1(A) + (1 - k') \cdot h^0(A)\right] + \left(1 - \frac{k}{k'}\right) \cdot h^0(A) \\
&= \lambda_\theta^k(A).
\end{aligned}$$

(To obtain the last two equalities, we use the null property of the experiment E_0). That establishes (ii) and completes the proof □

A.6 Proof of Theorem B6.

For the structure I_k, consider the signal $y_j \in C$. Its probability is

$$\sum_{i=1}^{n} [k \cdot \Lambda_{ij}^* \cdot p_i].$$

Its likelihood given the state θ_i is $k \cdot \Lambda^*_{ij}$. Given the signal y_j, the posterior probability of the state θ_t is

$$\frac{p_t \cdot k \cdot \Lambda^*_{tj}}{\sum_{i=1}^{n} \ k \cdot p_i \cdot \Lambda^*_{ij}} = \frac{p_t \Lambda^*_{tj}}{\sum_{i=1}^{n} \ p_i \cdot \Lambda^*_{ij}}.$$

So the vector of n state posteriors given the signal $y_j \in C$ is $\vec{\pi}^C(y_j) \in \Delta$, where

$$\vec{\pi}^C(y_j) = \frac{1}{\sum_{i=1}^{n} \ p_i \cdot \Lambda^*_{ij}} \cdot (p_1 \cdot \Lambda^*_{1j}, \ldots, p_n \cdot \Lambda^*_{nj}).$$

Similarly, consider the signal $y_\ell \in D$. Its probability is

$$\sum_{i=1}^{n} [(1-k) \cdot \Lambda_{i\ell} \cdot p_i].$$

Its likelihood given the state θ_i is $(1-k) \cdot \Lambda_{i\ell}$. The posterior probability of the state θ_t given the signal $y_\ell \in D$ is

$$\frac{p_t \cdot (1-k) \cdot \Lambda_{t\ell}}{\sum_{i=1}^{n} \ (1-k) \cdot p_i \cdot \Lambda_{i\ell}} = \frac{p_t \Lambda_{t\ell}}{\sum_{i=1}^{n} \ p_i \cdot \Lambda_{i\ell}}.$$

The vector of n state posteriors given $y_\ell \in D$ is $\vec{\pi}^D(y_\ell) \in \Delta$, where

$$\vec{\pi}^D(y_\ell) = \frac{1}{\sum_{i=1}^{n} \ p_i \cdot \Lambda_{i\ell}} \cdot (p_1 \cdot \Lambda_{1\ell}, \ldots, p_n \cdot \Lambda_{n\ell}).$$

The expected value of χ, over all the signals in the structure I_k, is

$$\begin{aligned}
\eta_\chi(k) \quad &= \quad \sum_{y_j \in C} [\text{ prob. of } y_j] \cdot \chi(\vec{\pi}^C(y_j)) + \sum_{y_\ell \in D} [\text{ prob. of } y_\ell] \cdot \chi(\vec{\pi}^D(y_\ell)) \\
&= \quad k \cdot \sum_{y_j \in C} [\sum_{i=1}^{n} \Lambda^*_{ij} \cdot p_i] \cdot \chi(\vec{\pi}^C(y_j)) + (1-k) \cdot \sum_{y_\ell \in D} [\sum_{i=1}^{n} \Lambda_{i\ell} \cdot p_i] \cdot \chi(\vec{\pi}^D(y_\ell)).
\end{aligned}$$

So, as claimed, η_χ is linear in k for all functions χ. Note that we did not need the subscript k on the symbols $\vec{\pi}_C, \vec{\pi}_D$, since the symbol k disappears when we calculate the posterior state probabilities. □

A.7 Proof of Theorem B7.

Conditions (i) and (ii) imply that

(†) $\breve{u}(q, F)$ has a unique maximizer in Q, namely $\hat{q}(F) = q^*(F)$.

We have to show that for any $\lambda \in [0,1]$ and any $F_1, F_2 \in \mathcal{F}$, we have

$$q^*(\lambda F_1 + (1-\lambda)F_2) \leq \lambda q^*(F_1) + (1-\lambda)q^*(F_2).$$

Consider the quantity

$$\xi\Big(\eta^*(\lambda q^*(F_1) + (1-\lambda)q^*(F_2)), \lambda F_1 + (1-\lambda)F_2\Big) \tag{1}$$

It will be convenient to use the shorthand symbol

$$q^\# \equiv \lambda q^*(F_1) + (1-\lambda)q^*(F_2).$$

The convexity of the function $\xi \circ \eta^*$ with respect to the pair (q, F) (assumed in (iii)) implies that (1) is less than or equal to

$$\lambda \cdot \xi\Big(\eta^*(q^*(F_1), F_1)\Big) + (1-\lambda) \cdot \xi\Big(\eta^*(q^*(F_2), F_2)\Big). \tag{2}$$

Using (†) and (iii), we have

$$\eta(q^*(F), F) = 0 \text{ and } \eta^*(q^*(F), F) = \eta^{**}(q^*(F), F)). \tag{3}$$

Therefore the expression (2) equals

$$\lambda \cdot \xi\Big(\eta^{**}(q^*(F_1), F_1)\Big) + (1-\lambda) \cdot \xi\Big(\eta^{**}(q^*(F_2), F_2)\Big). \tag{4}$$

But the concavity of the function $\xi \circ \eta^{**}$ with respect to the pair (q, F) (assumed in (iii)) implies that the expression in (4) is (using our shorthand symbol $q^\#$) less than or equal to

$$\xi\Big(\eta^{**}(q^\#, \lambda F_1 + (1-\lambda)F_2)\Big). \tag{5}$$

So we have shown that the expression in (1) is less than or equal to the expression in (5), i.e.,

$$\xi\Big(\eta^*(q^\#, \lambda F_1 + (1-\lambda)F_2)\Big) \leq \xi\Big(\eta^{**}(q^\#, \lambda F_1 + (1-\lambda)F_2)\Big). \tag{6}$$

But since (by (iii)) ξ is strictly increasing, (6) implies that

$$\eta^*(q^\#, \lambda F_1 + (1-\lambda)F_2) \leq \eta^{**}(q^\#, \lambda F_1 + (1-\lambda)F_2),$$

which implies, in turn, that

$$\eta(q^\#, \lambda F_1 + (1-\lambda)F_2) \leq 0. \tag{7}$$

But, by (3),

$$\eta(q^*(\lambda F_1 + (1-\lambda)F_2), \lambda F_1 + (1-\lambda)F_2) = 0.$$

Moreover, by assumption (i), the equation $\eta(q, \lambda F_1 + (1-\lambda)F_2) = 0$ has one and only one solution. So (by (†)) that unique solution is $q^*(\lambda F_1 + (1-\lambda)F_2)$. Then (7) and our assumption (ii) tell us that $\eta(q, \lambda F_1 + (1-\lambda)F_2)$ is zero at $q = q^*(\lambda F_1 + (1-\lambda)F_2)$, is positive for all smaller values of q, and is negative for all larger values. Hence, in view of (7), we have

$$q^*(\lambda F_1 + (1-\lambda)F_2) \leq q^{\#} = \lambda q^*(F_1) + (1-\lambda)q^*(F_2).$$

That completes the proof. □

A.8 Proof of Theorem N1

The proof uses two lemmata. *They are also used in the proofs of Theorems N2 and N3.* We start by proving the two lemmata.

Lemma C

Consider two sequences $\{V_n\}_{n\in\{1,2,\dots\}}$ and $\{W_n\}_{n\in\{1,2,\dots\}}$, where

$$V_n = \frac{\sum_{i=1}^n p_i}{\sum_{i=1}^n q_i};\ W_n = \frac{p_n}{q_n};\ p_i > 0, q_i > 0 \text{ for all } i \in \{1,2,\dots\}.$$

The following two statements hold:

[**1**] $W_1 < W_2 < \cdots < W_n < W_{n+1} < \cdots \Longrightarrow V_1 < V_2 < \cdots < V_n < V_{n+1} < \cdots$.

[**2**] $W_1 > W_2 > \cdots > W_n > W_{n+1} < \cdots \Longrightarrow V_1 > V_2 > \cdots > V_n > V_{n+1} > \cdots$.

Proof of Lemma C:

First note the following general fact:

(i) If $x > 0, y > 0, u > 0, v > 0$, then the following three statements are equivalent:

(a) $\frac{x}{y} > \frac{u}{v}$

(b) $\frac{x}{y} > \frac{x+u}{y+v}$

(c) $\frac{u}{v} < \frac{x+u}{y+v} < \frac{x}{y}$

In view of (i), we may claim:

(ii) for any $n \in \{1, 2, \dots\}$, $V_n < V_{n+1} \iff V_n < W_{n+1} \iff V_n < V_{n+1} < W_{n+1}$.

This follows from (i) and the following three statements, in which we identify the terms that play the role of "x", "y", "u", "v" in (i).

$$V_n = \frac{\overbrace{p_1 + \cdots + p_n}^{u}}{\underbrace{q_1 + \cdots + q_n}_{v}};\ V_{n+1} = \frac{\overbrace{[p_1 + \cdots + p_n]}^{x} + \overbrace{p_{n+1}}^{u}}{\underbrace{[q_1 + \cdots + q_n]}_{y} + \underbrace{q_{n+1}}_{v}};\ W_{n+1} = \frac{\overbrace{p_{n+1}}^{x}}{\underbrace{q_{n+1}}_{y}}.$$

We now prove [**1**] by induction. The statement holds for $n = 1$, since $V_1 = W_1 < W_2$ and hence $V_1 < V_2$ by (ii). Suppose that [**1**] holds for k, i.e., $V_k < V_{k+1}$. Then, by (ii), we have $V_k < V_{k+1} < W_{k+1}$, as well as $W_{k+1} < W_{k+2}$. Those two statements imply that $V_{k+1} < W_{k+2}$, which, by (ii), is equivalent to $V_{k+1} < V_{k+2}$. So [**2**] holds for $k+1$ and the induction is complete.

An analogous argument establishes [**2**]. *That concludes the proof of Lemma C.* □

Lemma D

Let (a, m) be a pair of real numbers such that

$$a > -1, \quad 1 + ma > 0.$$

Consider the function $\mu :\ (0, 1) \to \mathbb{R}$ on the positive reals defined by

$$\mu(x) = m \ln(1 - x^{1+a}) + (1 - m)\ln(1 - x) - \ln(1 - x^{1+am})), 0 < x < 1.$$

The function μ has the following properties, in each of three cases.

CASE 1: $m = 0$ or $m = 1$ or $a = 0$. In this case μ is constant.

CASE 2: $0 < m < 1$ and $a \neq 0$. In this case μ is strictly decreasing on $[0, 1)$.

CASE 3: $m > 1$ or $m < 0$ and $a \neq 0$. In this case μ is strictly increasing on $[0, 1)$.

Proof of Lemma D:

In Case 1 we have $\mu(x) = 0$ for all x, so μ is indeed constant.

We now proceed to the other two cases.

Structure of the proof for Cases 2 and 3.

The proof will exploit the fact that in these two cases we will be able to rewrite $\mu(x)$ in the form

$$\text{(L1)} \qquad k(x) = [\lambda f(x,u) + (1-\lambda) f(x,v)] - f(x, \lambda u + (1-\lambda)v),$$

for some $\lambda \in [0,1]$, some $u, v \in I\!R$, and some thrice differentiable function $f(x,y)$. Let f_x, f_{xyy} denote, respectively, the first derivative with respect to x and the third derivative (first with respect to x and then twice with respect to y). Note that

$$k'(x) = [\lambda f_x(x,u) + (1-\lambda) \cdot f_x(x,v)] - f_x(x, \lambda u + (1-\lambda) \cdot v).$$

Applying Jensen's inequality to k' for any fixed x, we obtain:

$$\text{(L2)} \qquad \text{If } f_{xyy} < 0, u \neq v, \text{ and } 0 < \lambda < 1, \text{ then } k \text{ is strictly decreasing .}$$

In the remaining steps of the proof we show that in each of the two cases, $\mu(x)$ can indeed be written in the form (L1) and that (L2) implies the claim which the Lemma makes for that case.

Step 1

The following will be a useful fact.

$$\text{(L3)} \qquad \text{for the function } \eta(x) = \frac{x}{e^x - 1}, x \in (0, \infty), \text{ we have } \eta''(x) > 0 \text{ for all } x > 0.$$

To show this, note that

$$\eta'(x) = \frac{e^x - 1 - xe^x}{(e^x - 1)^2}$$

and

$$\eta''(x) = \frac{-xe^x(e^x-1)^2 - (e^x - 1 - xe^x)2(e^x-1)e^x}{(e^x-1)^4} = \left[\frac{(e^x-1)e^x}{(e^x-1)^4}\right] \cdot \left[xe^x - 2e^x + 2 + x\right].$$

Since $x > 0$, we have $e^x - 1 > 0$. So the first term in large square brackets is positive. Now let $g(x)$ denote the second term in large square brackets, i.e., $g(x) \equiv xe^x - 2e^x + 2 + x$. We have $g'(x) = xe^x - e^x + 1$, and $g''(x) = xe^x$. Note that g'' is positive since $x > 0$. Hence $g'(x)$ is increasing in x. But $g'(0) = 0$. So, since $x > 0$, we have $g'(x) > g'(0) = 0$, i.e., g is increasing. Note that $g(0) = 0$ as well, so $g(x) > 0$ for positive x. That implies that $\eta''(x) > 0$ for $x > 0$.

Step 2

The following is another useful fact.

$$\text{(L4)} \qquad \begin{cases} \text{for the function } f(x,y) = \ln(1 - x^y), \text{ we have } f_{xyy} < 0 \text{ at any } (x,y) \text{ for} \\ \text{which } 0 < x < 1 \text{ and } y > 0. \end{cases}$$

That is the case since

$$f_x = \frac{-yx^{y-1}}{1-x^y} = \frac{-y/x}{x^{-y}-1} = \left(\frac{-y\ln x}{e^{-y\ln x}-1}\right)\cdot\left(\frac{1}{x\ln x}\right) = \eta(-y\ln x)\cdot\left(\frac{1}{x\ln x}\right),$$

where η is the function defined in (L3). Moreover,we have

$$f_{xyy} = \frac{\partial^2 f_x(x,y)}{\partial y^2} = \frac{\partial^2}{\partial y^2}\left[\eta((-y\ln x)\left(\frac{1}{x\ln x}\right)\right] = (-\ln x)^2\cdot\eta''(-y\ln x)\cdot\left(\frac{1}{x\ln x}\right).$$

We also have

(i) $\frac{1}{x\ln x} < 0$

(ii) $\eta''(t) > 0$ for positive t by (L4)

(iii) $-y\ln x > 0$ if $0 < x < 1, y > 0$.

We conclude that $f_{xyy} < 0$.

Step 3

Now consider Case 2.

We may write $\mu(x)$ in the form $k(x)$ of (L1) by choosing $f(x,y) = \ln(1-x^y)$, $\lambda = m \in (0,1)$, $u = 1+a$, and $v = 1$. We then have:

$$\mu(x) = m\ln(1-x^{1+a}) + (1-m)\ln(1-x^1) - \ln\left(1-x^{m(1+a)+(1-m)}\right) \text{ for } x \in (0,1).$$

By (L4) we have $f_{xyy} < 0$. Since $u \neq v$ (because $a \neq 0$), we can apply (L2). We conclude that μ is strictly decreasing on $(0,1)$. Moreover, since μ is continuous on $[0,1)$, it follwos that μ is, as claimed, strictly decreasing on $[0,1)$.

Step 4

Now consider a subcase of Case 3, namely

$$m > 1, a > -1, a \neq 0.$$

This time we may write $\mu(x)$ in the form

$$-m\cdot k(x),$$

where $k(x)$ has the form given in (L1). To obtain that form, we again choose $f(x,y) = \ln(1-x^y)$. We also choose $\lambda = \frac{m-1}{m}$, $u = 1+a$, and $v = 1+am$. We have

$$\mu(x) = -m\cdot k(x) = -m\cdot\left[\frac{m-1}{m}\cdot\ln\left(1-x^1\right) + \frac{1}{m}\cdot\ln\left(1-x^{1+am}\right) - \ln(1-x^{1+a})\right] \text{ for } x \in (0,1).$$

Using (L4) and (L2), we find that k is strictly decreasing and hence (since $-m < 0$), μ is strictly increasing on $(0, 1)$. Continuity of μ on $[0, 1)$ then implies that μ is strictly increasing on $[0, 1)$, as claimed for Case 3.

Step 5

Finally, we consider the remaining subcase of Case 3, namely

$$m < 0, a > -1, a \neq 0.$$

We write $\mu(x)$ in the form $-(1-m) \cdot k(x)$, where $k(x)$ again has the form given in (L1). To obtain that form, we again choose $f(x, y) = \ln(1 - x^y)$, $u = 1 + a$, and $v = 1 + am$. But now we choose $\lambda = \frac{-m}{1-m}$. We have

$$\mu(x) = -(1-m) \cdot k(x) =$$

$$-(1-m) \cdot \left[\frac{-m}{1-m} \cdot \ln\left(1 - x^{1+a}\right) + \frac{1}{1-m} \ln\left(1 - x^{1+am}\right) - \ln\left(1 - x^{1}\right)\right] \text{ for } x \in (0, 1).$$

Using (L4) and (L2), we find that k is strictly decreasing and hence (since $-(1-m) < 0$), μ is strictly increasing on $(0, 1)$. Continuity of μ on $[0, 1)$ then implies that μ is strictly increasing on $[0, 1)$, as claimed for Case 3.

That concludes the proof of Lemma D. □

Having established Lemma C and Lemma D, we now prove Theorem N1.

The state set is Θ is $[0, 1]$ and the prior G satisfies $G(\theta) = \theta^{\frac{1}{a}}$ for some $a > 0$. The statements we now obtain do not depend on a but only on m.

We have:

$$Y^n = \{y_1^n, \ldots, y_n^n\} = \{[x_0^n, x_1^n], (x_1^n, x_2^n], \ldots, (x_i^n, x_{i+1}^n], \ldots, (x_{n-2}^n, x_{n-1}^n], (x_{n-1}^n, x_n^n]\}$$

$$\text{with } x_0^n = u, x_n^n = v \text{ for all } n$$

and

(N1) $$G(x_i^n) - G(x_{i-1}^n) = \frac{1}{n} \text{ for all } i \in \{1, \ldots, n\}.$$

If the Producer receives the signal y_i^n, he knows that θ lies in $(x_i^n, x_{i+1}^n]$ and that $F_{y_i^n}^n$, the posterior cumulative distribution function on that interval, satisfies

(N2) $$F_{y_i^n}^n(\theta) = \frac{G(\theta) - (i-1)/n}{1/n} = n \cdot \left(G(\theta) - \frac{i-1}{n}\right).$$

We shall often omit the superscript n on the signals and on the posterior; the superscript will be understood.

In view of (N1) and (N2), the expected value of θ under the posterior F_{y_i}, for $i \in \{1, \ldots, n\}$, is

$$E_{F_{y_i}}\ \theta = \int_{x_{i-1}}^{x_i} \theta dF_{y_i}(\theta) = n \int_{x_{i-1}}^{x_i} \theta dG(\theta).$$

Defining the variable t to equal $G(\theta)$, we may rewrite this as

$$\text{(N3)} \qquad E_{F_{y_i}}\ \theta = n \int_{\frac{i-1}{n}}^{\frac{i}{n}} G^{-1}(t)dt.$$

The proof proceeds in four steps.

Step 1

If $m = 0$, then $\phi = 1$ and $A_n = 1$ for all n.

If $m = 1$, then $\phi(t) = t$ and A_n equals the expected value, over all signals in y in Y^n, of the posterior means $E_{F_y^n}\ \theta$. For all n, that expected value equals $E_G\ \theta$, the mean of the prior.

So [**3**] is established.

Henceforth we assume that $m \neq 0$ and $m \neq 1$.

Step 2

Using (N3) above and the fact that $G^{-1}(t) = t^a$, we have

$$E_{F_{y_i}}(\theta) = n \int_{\frac{i-1}{n}}^{\frac{i}{n}} G^{-1}(t)dt = n \int_{\frac{i-1}{n}}^{\frac{i}{n}} t^a dt = \frac{n}{1+a} \cdot \left[\left(\frac{i}{n}\right)^{1+a} - \left(\frac{i-1}{n}\right)^{1+a} \right].$$

Hence

$$A_n = \frac{1}{n} \sum_{i=1}^{n} \phi\left(E_{F_{y_i}}\ \theta\right) = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{n}{1+a} \left[\left(\frac{i}{n}\right)^{1+a} - \left(\frac{i-1}{n}\right)^{1+a} \right] \right)^m = \frac{\sum_{i=1}^{n} \left[i^{1+a} - (i-1)^{1+a}\right]^m}{(1+a)^m \cdot n^{1+am}}.$$

The numerator of this fraction is strictly increasing in n , since each term $\left[i^{1+a} - (i-1)^{1+a}\right]^m$ is positive and we add one more such term when we move from n to $n+1$. If $1 + am \leq 0$, i.e., if $m \leq -\frac{1}{a} < 0$, the denominator is weakly decreasing in n. We conclude that

$$\text{(N4)} \qquad \text{if } m \leq -\frac{1}{a} \text{ then } A_n \text{ is strictly increasing in } n.$$

Step 3

Now consider the case $m > -\frac{1}{a}$ or $1 + am > 0$. For every integer $n \geq 1$ define

$$B_n \equiv \frac{[n^{1+a} - (n-1)^{1+a}]^m}{n^{1+am} - (n-1)^{1+am}}.$$

We shall use the following Claim, which follows from Lemma C:

Claim 1

Suppose that $1 + am > 0$. Then the following statements hold:
If B_n is strictly increasing in n then so is A_n. If B_n is strictly decreasing in n then so is A_n.

We note that

(N5) $$B_n = \frac{[1 - (\frac{n-1}{n})^{1+a}]^m}{1 - (\frac{n-1}{n})^{1+am}} \cdot \frac{n^{m(1+a)}}{n^{1+am}} = \frac{[1 - (\frac{n-1}{n})^{1+a}]^m}{1 - (\frac{n-1}{n})^{1+am}} \cdot \frac{1}{n^{1-m}}.$$

Now define

$$x(n) \equiv \frac{n-1}{n}.$$

Note that x is increasing in n and that

(N6) $$1 - x(n) = \frac{1}{n}.$$

We abbreviate $x(n)$ as x. The variable x takes values in $[0, 1)$.

Using (N6) and the last term in (N5), we may write

(N7) $$B_n = H(x) \equiv \frac{(1 - x^{1+a})^m}{1 - x^{1+am}} \cdot (1-x)^{1-m}.$$

It will be useful to take logarithms on both sides of (N7). We observe that since logarithm is strictly increasing, we have:

> $H(x)$ is strictly increasing (strictly decreasing) in x if $\ln H(x)$ is strictly increasing (strictly decreasing) in x.

Since $x = (n-1)/n$ is an increasing function of n, Claim 1 and (N7) imply the following:

(N8) $$\begin{cases} \text{Suppose } 1 + am > 0. \text{ If } \ln H(x) \text{ is strictly increasing} \\ \text{(strictly decreasing) in } x \text{ at all } x \in [0, 1[, \text{ then } A_n \text{ is} \\ \text{strictly increasing (strictly decreasing) in } n \end{cases}$$

Step 4

We now state a second claim. It follows from Lemma D and the fact that $\ln H(x)$ is the same as the function $\mu(x)$ of Lemma D. (Note that Lemma D permits a to be negative and greater than -1, but for the present theorem we only need $a > 0$).

<u>**Claim 2**</u>

Suppose $1 + am > 0$. Then for all $a > 0$:

(1) $\ln H(x)$ is strictly increasing at all $x \in [0, 1)$ if $m > 1$ or $m < 0$.

(2) $\ln H(x)$ is strictly decreasing at all $x \in [0, 1)$ if $0 < m < 1$.

We now assemble what we have found.

Conclusion [**3**] of the Theorem deals with the cases $m = 1$ and $m = 0$ and was established in Step 1.

Conclusion [**1**] of the Theorem deals with the cases $m > 1$ and $m < 0$. The subcase $m \leq -1/a$, or $am + 1 \leq 0$, is covered by (N4). The subcase $-1/a < m < 0$, and the case $m > 1$ are covered by (1) of Claim 2 combined with (N8).

Conclusion [**2**] of the Theorem deals with the cases $0 < m < 1$ and is established by (**2**) of Claim 2 combined with (N8).

That concludes the proof of Theorem N1. □

A.9 Proof of Theorem N2

The proof repeats the pattern of the proof of Theorem N1.

Note first that

$$G^{-1}(t) = k(1-t)^{-1/\delta}$$

and define

$$a \equiv -\frac{1}{\delta}.$$

Using (N3), we have

$$E_{F_{y_i}}\ \theta = n \int_{\frac{i-1}{n}}^{\frac{i}{n}} G^{-1}(t)\ dt = nk \int_{\frac{i-1}{n}}^{\frac{i}{n}} (1-t)^a dt.$$

Now define the variable $u \equiv 1-t$. We obtain

$$E_{F_{y_i}}\ \theta = -nk\int_{\frac{n-i+1}{n}}^{\frac{n-i}{n}} u^a\ du = nk\int_{\frac{n-i}{n}}^{\frac{n-i+1}{n}} u^a du = \frac{nk}{1+a}\cdot\left[\left(\frac{n-i+1}{n}\right)^{1+a} - \left(\frac{n-i}{n}\right)^{1+a}\right].$$

So

$$A_n = \frac{1}{n}\sum_{i=1}^{n}\phi\left(E_{F_{y_i}^n}\ \theta\right) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{nk}{1+a}\cdot\left(\left(\frac{n-i+1}{n}\right)^{1+a} - \left(\frac{n-i}{n}\right)^{1+a}\right)\right]^m$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left[\frac{nk}{1+a}\cdot\left(\left(\frac{i}{n}\right)^{1+a} - \left(\frac{i-1}{n}\right)^{1+a}\right)\right]^m = \frac{k^m}{(1+a)^m}\cdot\frac{\sum_{i=1}^{n}[i^{1+a}-(i-1)^{1+a}]^m}{n^{1+am}}.$$

The final expression is the same as the expression for A_n in the proof of Theorem N1. Now, however, we have $a \in (-1,0)$, while in Theorem N1 we had $a > 0$. Inspecting the proof of Theorem N1, we see that all steps of the proof can be repeated if a is required to lie in $(-1,0)$. (See, in particular the remark about Lemma B at the start of Step 4).

That concludes the proof. □

A.10 Proof of Theorem N3

The statement [**3**] is easily established (the argument is analogous to the proof of [**3**] in Theorem N1).

The probability of the interval defined by the signal y_i is

$$G(\frac{i}{n}) - G\left(\frac{i-1}{n}\right).$$

So for the posterior cumulative distribution function $F_{y_i}^n$ we have

$$F_{y_i}^n(\theta) = \frac{G(\theta) - G\left(\frac{i-1}{n}\right)}{G\left(\frac{i}{n}\right) - G\left(\frac{i-1}{n}\right)}.$$

Its mean is

$$E_{F_{y_i}^n}\ \theta = \int_{\frac{i-1}{n}}^{\frac{i}{n}} \theta\ dF_{y_i}^n(\theta).$$

It will be useful to define

$$b \equiv \frac{1}{d}.$$

Since $G(\theta) = \theta^b$, we have

$$E_{F_{y_i}^n}\ \theta = \int_{\frac{i-1}{n}}^{\frac{i}{n}} \theta\cdot\frac{dG(\theta)}{G\left(\frac{i}{n}\right) - G\left(\frac{i-1}{n}\right)} = \frac{\int_{\frac{i-1}{n}}^{\frac{i}{n}}\ \theta\cdot b\cdot\theta^{b-1}\ d\theta}{G\left(\frac{i}{n}\right) - G\left(\frac{i-1}{n}\right)} = \frac{b}{1+b}\cdot\left(\frac{\left(\frac{i}{n}\right)^{1+b} - \left(\frac{i-1}{n}\right)^{1+b}}{\left(\frac{i}{n}\right)^{b} - \left(\frac{i-1}{n}\right)^{b}}\right).$$

Hence

$$A_n = \sum_{i=1}^{n} (\text{ prob. of the } y_i \textbf{ interval }) \cdot \left[E_{F^n_{y_i}}\ \theta\right]^m$$

$$= \sum_{i=1}^{n}\left[\left(\frac{i}{n}\right)^b - \left(\frac{i-1}{n}\right)^b\right]\cdot\left[\frac{b}{1+b}\cdot\frac{\left(\frac{i}{n}\right)^{1+b}-\left(\frac{i-1}{n}\right)^{1+b}}{\left(\frac{i}{n}\right)^{b}-\left(\frac{i-1}{n}\right)^{b}}\right]^m = \sum_{i=1}^{n}\left(\frac{b}{1+b}\right)^m\cdot\frac{\left[\left(\frac{i}{n}\right)^{1+b}-\left(\frac{i-1}{n}\right)^{1+b}\right]^m}{\left[\ \left(\frac{i}{n}\right)^{b}-\left(\frac{i-1}{n}\right)^{b}\ \right]^{m-1}}$$

$$= \sum_{i=1}^{n}\left(\frac{b}{1+b}\right)^m\cdot\frac{\left[i^{1+b}-(i-1)^{1+b}\right]^m}{\left[i^b-(i-1)^b\right]^{m-1}}\cdot\frac{1}{n^{m\cdot(1+b)-(m-1)\cdot b}} = \left(\frac{b}{1+b}\right)^m\cdot\frac{\left(\ \sum_{i=1}^{n}\frac{\left[i^{1+b}-(i-1)^{1+b}\right]^m}{\left[i^b-(i-1)^b\right]^{m-1}}\right)}{n^{m+b}}.$$

The following is easily verified:

(1) If $m + b \leq 0$, then A_n is strictly increasing in n.

Accordingly we confine attention, henceforth, to the case

$$m + b > 0.$$

Now define

$$B_n \equiv \frac{\left(\frac{\left[n^{1+b}-(n-1)^{1+b}\right]^m}{\left[n^b-(n-1)^b\right]^{m-1}}\right)}{n^{m+b}-(n-1)^{m+b}}, n = 0, 1, 2, \ldots.$$

Note that $\left(\frac{b}{1+b}\right)^m > 0$ and consider the term that follows $\left(\frac{b}{1+b}\right)^m$ in the final expression for A_n. An argument that uses Lemma C in the proof of theorem N1, and is analogous to the argument establishing Claim 1 in that proof, tells us that

(2) A_n is strictly increasing (strictly decreasing) in n if B_n is strictly increasing (strictly decreasing) in n.

Next, for $n \in \{1, 2, 3, \ldots,\}$, define

$$z(n) \equiv \frac{n-1}{n}$$

and note that $z \in [0, 1)$ and that z is strictly increasing in n. We have

$$B_n = \frac{n^{(1+b)\cdot m}\cdot\left(1-\left(\frac{n-1}{n}\right)^{1+b}\right)^m}{n^{b\cdot(m-1)}\cdot\left(1-\left(\frac{n-1}{n}\right)^{b}\right)^{m-1}}\cdot\frac{1}{n^{m+b}\cdot\left(1-\left(\frac{n-1}{n}\right)^{m+b}\right)}$$

$$= n^{(1+b)\cdot m-b\cdot(m-1)-(m+b)}\cdot\frac{\left(1-z^{1+b}\right)^m}{\left(1-z^b\right)^{m-1}\cdot\left(1-z^{m+b}\right)} = \frac{\left(1-z^{1+b}\right)^m}{\left(1-z^b\right)^{m-1}\cdot\left(1-z^{m+b}\right)}.$$

Call the final term $H(z)$. For $z \in [0,1)$ we now consider

$$\ln H(z) = m \ln\left(1 - z^{1+b}\right) - (m-1)\ln\left(1 - z^{b}\right) - \ln\left(1 - z^{m+b}\right).$$

If $\ln H(z)$ rises (falls) when z increases, then so does $H(z)$ and thus (since z is strictly increasing in n, $H(z(n)) = B_n$ rises (falls) when n increases. We shall use Lemma B in the proof of Theorem N1 to investigate the behavior of $\ln H(z)$ as z increases in $[0,1)$. To do so, the following changes in variables are helpful: Define

$$x \equiv z^b;\ a \equiv \frac{1}{b}.$$

Then z^{1+b} becomes x^{1+a} and z^{m+b} becomes x^{1+am}. Thus $\ln H(z)$ becomes precisely the expression $\mu(x)$ considered in Lemma D. Applying Lemma D, and recalling that the case $m+b<0$ was already dealt with in (1), we obtain the following:

<u>**Case (a)** $m > 1$ **or** $-b < m < 0$**.**</u>

This fits Case (3) of Lemma D. The expression $\mu(x)$ rises when $x = z^b$ rises. We conclude that

(3) if $m > 1$ or $m < 0$, then B_n rises when n increases.

<u>**Case (b)** $0 < m < 1$**.**</u>

This is Case (2) of Lemma D. The expression $\mu(x)$ falls when $x = z^b$ rises. We conclude that

(4) if $0 < m < 1$, then B_n falls when n increases.

Combining (3),(4), with (1),(2) completes the proof. □

B Constructing the counterexample discussed in Section 3.2.2

The counterexample is constructed by choosing a prior distribution G and six adjoining equal-probability sub-intervals of $[0,1]$, denoted $I_1,\ldots,I_6$, where

$$I_v = (a_v, b_v).$$

The two-interval structure has the equal-probability intervals $I_1 \cup I_2 \cup I_3$ and $I_4 \cup I_5 \cup I_6$. The three-interval structure has the equal-probability intervals $I_1 \cup I_2$, $I_3 \cup I_4$, and $I_5 \cup I_6$. Let m_1, m_2 denote the two conditional means for the two intervals of the first structure, and let n_1, n_2, n_3

denote the three conditional means in the second structure. Then for the first structure the average best quantity is

$$\frac{1}{2} \cdot \left((m_1)^2 + (m_2)^2\right)$$

and the average maximal expected payoff is

$$\frac{1}{2} \cdot \left(\frac{1}{3} \cdot (m_1)^3 + \frac{1}{3} \cdot (m_2)^3\right) = \frac{1}{6} \cdot \left((m_1)^3 + (m_2)^3\right).$$

For the second structure the average best quantity is

$$\frac{1}{3} \cdot \left((n_1)^2 + (n_2)^2 + (n_3)^2\right)$$

and the average maximal expected payoff is

$$\frac{1}{3} \cdot \left(\frac{1}{3} \cdot (n_1)^3 + \frac{1}{3} \cdot (n_2)^3 + \frac{1}{3} \cdot (n_3)^3\right) = \frac{1}{9} \cdot \left((n_1)^3 + (n_2)^3 + (n_3)^3\right).$$

In the example we seek, average best payoff rises when we go from two intervals to three but average best quantity falls, i.e., we have

$$\frac{1}{9} \cdot \left((n_1)^3 + (n_2)^3 + (n_3)^3\right) > \frac{1}{6} \cdot \left((m_1)^3 + (m_2)^3\right),$$

or equivalently

$$(C1) \qquad \frac{1}{3} \cdot \left((n_1)^3 + (n_2)^3 + (n_3)^3\right) > \frac{1}{2} \cdot \left((m_1)^3 + (m_2)^3\right),$$

and at the same time

$$(C2) \qquad \frac{1}{3} \cdot \left((n_1)^2 + (n_2)^2 + (n_3)^2\right) < \frac{1}{2} \cdot \left((m_1)^2 + (m_2)^2\right).$$

We select six numbers $k_1, \ldots, k_6$, where

$$k_v \in (a_v, b_v), v \in \{1, 2, \ldots, 6\}.$$

We can then construct a prior G on $\Theta = [0, 1]$, with a density function g having the following properties:

- $g(a_v) = g(b_v) = 0$.
- The mass on each interval (a_v, b_v) equals $\frac{1}{6}$, i.e.,

$$(C3) \qquad \text{for every } v \in \{1, \ldots, 6\}, \text{ we have } \int_{a_v}^{b_v} f(r)dr = \frac{1}{6}.$$

- Each k_v equals the conditional expected value of θ, given that $\theta \in (a_v, b_v)$, i.e.,

$$(C4) \qquad \text{for every } v \in \{1, \ldots, 6\}, \text{ we have } k_v = \frac{\int_{a_v}^{b_v} r \cdot g(r)dr}{\int_{a_v}^{b_v} g(r)dr}.$$

One way to construct such a density g is as follows. For each v, we select a sub-interval $(a_b^*, b_v^*) \subseteq (a_v, b_v)$ and we choose f so that f and each (a_v^*, b_v^*) meet the following conditions:

(i) $k_v \in (a_v^*, b_v^*)$.

(ii) $g(\theta) = 0$ for $a_v \leq \theta \leq a_v^*$ and for $b_v \geq \theta \geq b_v^*$;

(iii) for $\theta \in (a_b^*, b_v^*)$, the graph of $g(\theta)$ (placing $[0, 1]$ on the horizontal axis) defines a triangle. Its base is (a_v^*, b_v^*) and its upper vertex is the point (ℓ_v, h_v).

(iv) $\frac{1}{2} \cdot (b_v^* - a_v^*) \cdot h_v = \frac{1}{6}$.

(v) $\dfrac{a_v^* + b_v^* + \ell_v}{3} = k_v$.

Now for each v, let us select $a_v^*, b_v^*, \ell_v, h_v$ by choosing an $\epsilon > 0$ and defining

$$a_v^* = k_v - \epsilon, b_v^* = k_v + \epsilon, \ell_v = k_v, h_v = \frac{1}{6\epsilon}.$$

Then for sufficiently small ϵ, (i) is met. For all ϵ, both (iv) and (v) are met. Conditions (iv) and (ii) guarantee that the mass on (a_v, b_v) is indeed $\frac{1}{6}$, i.e., that (C3) is indeed satisfied. In view of (C3), the total mass for the f we have constructed equals one, so f is indeed a density function. Conditions (iii) and (v) guarantee (using the formula for the mean of a triangular distribution) that the conditional expected value of θ given that $\theta \in (a_v^*, b_v^*)$ (which, in view of (ii), equals the expected value of θ given that $\theta \in (a_v, b_v)$) indeed equals k_v, i.e., (C4) is indeed satisfied.

The graph of our density f coincides with the θ-axis outside the intervals (a_v^*, b_v^*). In each interval (a_v^*, b_v^*), the graph defines a triangle whose base is that interval.

To summarize, we have shown the following:

> Suppose we have found six adjacent intervals $(a_v, b_v) \subset [0, 1]$ and six numbers $k_1, \ldots, k_6$ such that
>
> - $k_v \in (a_v, b_v), v \in \{1, \ldots, 6\}$

- (C1) and (C2) are satisfied by m_1, m_2, n_1, n_2, n_3, where

$$m_1 = \frac{1}{3} \cdot (k_1 + k_2 + k_3), m_2 = \frac{1}{3} \cdot (k_4 + k_5 + k_6)$$

and

$$n_1 = \frac{1}{2} \cdot (k_1 + k_2), n_2 = \frac{1}{2} \cdot (k_3 + k_4), n_3 = \frac{1}{2} \cdot (k_5 + k_6).$$

Then there exists a prior on $[0, 1]$ such that

- The intervals (a_v, b_v) have equal probability.
- The expected value of θ given that $\theta \in I_v$ equals k_v
- If we go from a two-signal structure whose signals identify the equal-probability intervals $I_1 \cup I_2 \cup I_3, I_4 \cup I_5 \cup I_6$ to a three-signal structure whose signals identify the equal-probability intervals $I_1 \cup I_2, I_3 \cup I_4, I_5 \cup I_6$, then expected best payoff (for the payoff function $q\theta - \frac{2}{3} \cdot q^{3/2}$) rises but average best quantity falls.

Note that in searching for an example we need not confine attention to six intervals (a_v, b_v) having equal *width*. It turns out, nevertheless, that we obtain the example we seek if we let the six adjacent intervals (a_v, b_v) be

$$(0, 1/6), (1/6, 1/3), (1/3, 1/2), (1/2, 2/3), (2/3, 5/6), (5/6, 1),$$

and we choose the following six numbers $k_1, \ldots, k_6$:

$$0.136985,\ 0.177125,\ 0.377946,\ 0.662507,\ 0.704948,\ 0.905665.$$

For each v, we have $k_v \in (a_v, b_v)$. We obtain

$$m_1 = 0.247168;\ m_2 = 0.804017$$

and

$$n_1 = 0.196429;\ n_2 = 0.503433;\ n_3 = 0.876916.$$

Next we obtain

$$\frac{(m_1)^3 + (m_2)^3}{2} - \frac{(n_1)^3 + (n_2)^3 + (n_3)^3}{3} = -0.00240854 < 0,$$

which is equivalent to (C1), and

$$\frac{(m_1)^2 + (m_2)^2}{2} - \frac{(n_1)^2 + (n_2)^2 + (n_3)^2}{3} = 0.0000976841 > 0,$$

which is equivalent to (C2).

So the example we sought has been found.

REFERENCES

Acemoglu, D., Autor, D. (2011), "Skills, Tasks, and Technologies: Implications for Employment and Earnings", Handbook of Labor Economics, Volume 4b, Elsevier.

Amershi, A. (1988), "Blackwell Informativeness and Sufficient Statistics with Applications to Financial Markets and Multiperson Agencies", in Feltham, G, Amershi, A., and Ziemba, W., *Economic Analysis of information and Contracts*, Boston: Kluwer Academic Publishers.

Athey, S. and Levin, J. (2001), "The Value of Information in Monotone Decision Problems", working paper.

Autor, D., Levy, F., Murnane, R. (2003) "The Skill Content of Recent Technological Change: An Empirical Exploration, *Quarterly Journal of Economics*, 118, 1279-1333.

Blackwell, D. (1951), "Comparison of Experiments", *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (J. Neyman, ed.), 93-102.

Blackwell, D. (1953), "Equivalent Comparison of Experiments", *Annals of Mathematical Statistics*, **24**, 265-273.

Bresnahan, T., E. Brynjolfsson, L. Hitt (2002), "Information Technology, Workplace Organization, and the Demand for Skilled Labor", *Quarterly Journal of Economics*, 118, 339-376.

Cover, B., Jones, J., and Watson, A. (2011), "Science, Technology, Engineering, and Mathematics (STEM) Occupations: A Visual essay", *Monthly Labor Review*, May.

Crémer, J. (1982), "A Simple Proof of Blackwell's Comparison of Experiments Theorem", *Journal of Economic Theory*, 27, 439-443.

Green, J., and Stokey, N. (2007), "A Two-person Game of Information Transmission", *Journal of economic theory*, 135, 90-104.

Holthausen, H. (1976), "Input Choices and Uncertain Demand", *American Economic Review*, Vol. 66, No. 1.

Hubbard, T. (2000), "The Demand for Monitoring Technology: The Case of Trucking, *Quarterly Journal of Economics*, 115, 533-560.

Hubbard, T. (2003), "Information, Decision, and Productivity: On-board Computers and Capacity Utilization in Trucking", *American Economic Review*, 93, 4, 1328-1353.

Johnson, P. and Myatt, D. (2006), "On the Simple Economics of Advertising, Marketing, and Product Design", *American Economic Review*, 96 (3).

Jorgenson, D., M. Ho, K. Shiroh (2005), *Productivity, Volume 3: Information Technology and the American Growth Resurgence*, MIT Press.

Kretschmer, T. (2012), "Information and Communication Technologies and Productivity Growth: A Survey of the Literature", *OECD Digital Economy Papers, No. 195*, Paris, OECD Publishing.

Le Cam, L. (1964), "Sufficiency and Approximate Sufficiency", *Ann. Math. Stat.* 35, 1419-1455.

Le Cam, L. (1996), "Comparison of Experiments: A Short Review", in *Statistics, Probability, and Game Theory: Papers in Honor of David Blackwell*, Institute of Mathematical Statistics Lecture Notes Monograph Series, Volume 30.

Lee, M. and Mather M. (2008), "US Labor Force Trends, *Population Bulletin*, Vol 63, No. 2.

Lehman E. (1955), "Ordered Families of Distributions", Annals Math. Stat., Vol. 26, No. 3.

Lehman E. (1988), "Comparing Location Experiments", Annals Math. Stat., Vol. 59, No. 2, 521-533.

Leland. H., (1972), "Theory of the Firm Facing Uncertain Demand", *American Economic Review*, 62.

Marschak, J., Miyasawa, K. (1968), "Economic Comparability of Information Systems", *International Economic Review*, **9**, 137-174.

Marschak, T., Shanthikumar J.G., and Zhou, J. (October 2013), "A Newsvendor Who Chooses Informational Effort", working paper.

Matthews, H. and Stine, C. (2009), "The US Science and Technology Workforce', CRS Report RL43539, Congressional Research Service, June.

McGuire, C.B. (1972), "Comparison of Information Structures", in C.B. McGuire and R. Radner (eds.), *Decision and Organization: A Volume in Honor of Jacob Marschak*, North-Holland.

Milgrom, P. and Shannon, C. (1994), "Monotone Comparative Statics", *Econometrica*, **62**, 157-180.

Ponssard, J.P. (1975), "A Note on Information Value Theory for Experiments Defined in Extensive Form", *Management Science*, 22 (4).

Rajan, M.V. and Saouma, R. (2006), "Optimal Information Asymmetry", *Accounting Review*, 81 (3).

Sandmo, A. (1971), "On the Theory of the Competitive Firm Under Price Uncertainty", *American Economic Review*, Vol 61, No. 1.

Shaked, M. and Shanthikumar, J.G. (2007), *Stochastic Orders*, New York: Springer.

Strassen, V. (1965), "The Existence of Probability Measures with Given Marginals", *Annals of Mathematical Statistics*, **36**, 29, 423-439.

Torgersen, E.N. (1976), "Comparison of Statistical Experiments", *Scandinavian Journal of Staistics*, 3, 186-208.

Torgersen, E.N. (1991), *Comparison of Statistical Experiments*, Cambridge University press.

Zhang, Zhixiang (2008), "Comparison of Information Structures with Infinite States of Nature", dissertation submitted to Johns Hopkins University (Department of Economics).