MIT LIBRARIES DUPL
3 9080 00665066 4

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

HD28
.M414
no. 3254-
91

DEWEY

MAR 14 1991

WORKING PAPER
ALFRED P. SLOAN SCHOOL OF MANAGEMENT

Johnny Carson vs. the Smothers Brothers

Monolog vs. Dialog in Costly Bilateral Communication

Fred Kofman
Massachusetts Institute of Technology

Jim Ratliff
University of California, Berkeley

WP# 3254-91-EFA

March 1991

MASSACHUSETTS
INSTITUTE OF TECHNOLOGY
50 MEMORIAL DRIVE
CAMBRIDGE, MASSACHUSETTS 02139

Johnny Carson vs. the Smothers Brothers

Monolog vs. Dialog in Costly Bilateral Communication

Fred Kofman
Massachusetts Institute of Technology

Jim Ratliff
University of California, Berkeley

WP# 3254-91-EFA March 1991

Dewey
M.I.T. LIBRARIES
MAR 1 4 1991
RECEIVED

Johnny Carson vs. the Smothers Brothers

Monolog vs. Dialog in Costly Bilateral Communication

Fred Kofman
Sloan School of Management
Massachusetts Institute of Technology

Jim Ratliff
Department of Economics
University of California, Berkeley

February 27, 1991

This research was supported by National Science Foundation Grant IRI-8902813 and by the Sloan Foundation.

1. Introduction

Members of organizations communicate with each other in order to coordinate their actions in pursuit of a common goal. Bilateral information exchanges can be broadly classified as monologs or dialogs. A typical monolog would be a published forecast of market conditions in response to which a manager chooses an inventory level. A typical dialog could be a question-and-answer session; e.g. the chief strategist describes a new product, the marketing prognosticator announces a sales estimate based on that description, and then the strategist decides whether to introduce the product.

Our common sense and experience tell us that dialog, if not in some sense dominant, is certainly a ubiquitous mode of communication in the real world. It is provocative, then, to compare this observed dialog-rich communication with the type of discourse manifested in traditional economic analysis. In most models with asymmetric information an agent is called upon to report her private information to a principal. Consideration is usually restricted to mechanisms in which the agent fully and truthfully reveals the entirety of her private information (i.e. she declares her "type").[1] In other words these are models of a monolog world. (There is no need to ask a question of someone who is planning to tell you everything she knows, because you know that you will receive the answer to every question which she could possibly answer.)

Why do we observe dialog in the world but not in our models? Theorists invoke the Revelation Principle in order to claim that no generality is lost when they focus upon tell-everything monolog mechanisms. This exploitation of the Revelation Principle is only valid, however, when communication is costless.[2] When communication is costly, full revelation of type may be so expensive as to be suboptimal, and efficiency in communication becomes a relevant criterion for the mechanism designer. This raises new questions of who should say what to whom and when and thereby introduces the possibility of observing dialog.[3]

The costs of communication is a theoretical lacuna which begs to be closed. As Tirole [1988: 49] has pointed out: "Neoclassical theory pays only lip service to the issue of communication. Information flows between members of an organization are limited only because of incentive compatibility ... However, even well-intentioned members of an organization ... may have trouble communicating all the information they possess to their relevant co-members, because it is too time consuming or because the information is hard to 'codify' to make it understandable to its receivers. Thus, decisions that would be profit maximizing under full communication will not be made under imperfect communication." Arrow [1974: 5] advocates that before we can rigorously evaluate the market mechanism's claim to superior

1 In some circumstances efficiency can be achieved without full revelation. See Groves and Ledyard [1977], Drèze and de la Vallée Poussin [1971], and Malinvaud [1971].

2 McAfee and McMillan [1988] extend the Revelation Principle to a case in which the principal bears a communication cost but the agents do not.

3 Even more generally, costly communication has the potential to explain the existence of hierarchical organizational structures. (One interpretation of the Revelation Principle holds that the outcome of a decentralized organization can be replicated by a centralized two-tier structure; therefore hierarchy can never be strictly preferred.) See Melumad, Mookherjee, and Reichelstein [1989] for work in this direction which adopts a message space dimensionality perspective on communication cost.

informational economy we must "add to our usual economic calculations an appropriate measure of the costs of information gathering and transmission."

Communication costs lead to bounded rationality because they impose cognitive limitations—via knowledge restrictions—upon decisionmakers. Limited communication is similar formally to other sources of bounded rationality, for example limited memory. Dow [1991] studies an agent searching for a low price and asks how the agent can make optimal use of her limited memory for observed prices. Clearly this is equivalent to a team[1] communication problem with two partners, Past and Present, where the design decision is how Past should use limited communication resources to inform Present about the price Past observed.[2]

When communication is free and truthful revelation is assured, all relevant information can be transmitted to a specified decisionmaker. However, when communication costs inhibit a comprehensive information exchange, the initial informational asymmetries are not completely obliterated; by the end of the conversation some agents still know some things which others do not know. Some agent may be more qualified to make the choice-of-action decision, and therefore the designer must determine to whom the decision would be optimally delegated and under what circumstances.

In this paper we posit the existence of communication costs in a bilateral team context. Our task is to determine the exact process by which private information should be shared and decisionmaking authority should be delegated. We will see that indeed there exist cases in which it would be suboptimal for one party to engage in a monolog with the other; i.e. a dialog would achieve a higher performance standard for a fixed communication cost.

We are intrigued by three features of the optimal information sharing mechanisms we have encountered during our research. First, they can have an unpredictable life of their own. To guarantee efficiency it is not sufficient that we merely enrich the designer's toolkit to include a simplistic capability of alternating question and answer. A rigid, fixed-sequence dialog—which dictates that the microphone change hands in a particular, predetermined order and prescribes ahead of time a particular agent as the ultimate decisionmaker—can be suboptimal just as a rigid monolog can be. Instead, the designer must supply a communication *algorithm*, according to which the identity of the speaker at any stage and the identity of the ultimate decisionmaker are determined endogenously by the particular realization of the agents' private data. Just as a software engineer cannot predict precisely what output will result from a computer program without first knowing its input data, the mechanism designer cannot—in ignorance of the agents' private information—predict the sequence of speakers in the conversation or the identity of the decisionmaker.

Secondly, it can be strictly better if occasionally neither party speaks—letting one of them choose an action in complete ignorance of the other's private information—*even when they have already paid for the right to communicate* and *when the choice of action could be improved by more precise knowledge*

[1] A team in the sense of Marschak and Radner [1972] is a group whose members have only common interests.

[2] This communication is clearly limited to a monolog in which Past speaks to Present. The reverse direction would be a case of paranormal fortunetelling rather than bounded rationality!

of that private information.[1] The intuition for this counterintuitive result is that a deliberate silence in some circumstances increases the informativeness of communication in the remaining circumstances. A loss in decisionmaking quality due to remaining silent in a circumstance in which communication would be relatively ineffective anyway can be traded-off against a gain generated by the resulting increased informativeness in cases where communication is more crucial.[2]

A third interesting feature of these optimal communication algorithms concerns who dominates the conversation. We will see that the identity of the agent who should do the most talking can depend on the size of the communication budget. If you cannot pay for much talking, perhaps one agent is the preferred lecturer. If you have more to spend, it can be optimal to hear instead from the other.

Although we show that dialog is sometimes necessary for efficiency, we do establish sufficient conditions under which monolog is efficient. Mathematically this takes the form of an additive separability requirement. An interpretation of this result is that monolog will be sufficient when the error in the final result—generated by a given misestimate by the decisionmaker of her partner's private information—is independent of the decisionmaker's private information. For monolog efficiency, then, it is sufficient that the decision problem have a structure in which the partners' private data do not interact in this well-defined sense.

The problem

In order to focus on the analytical issues raised by costly communication we will abstract away from incentive issues and study the following problem:[3] Consider a team of two privately informed members. The designer's task is to construct a mechanism through which the team members share their private information with one another in order that one of them will make a decision which would optimally depend on both of their privately known data. (The designer does not know at the time of design what each agent's private datum will be at the time the mechanism is implemented.)

Our communication cost measure is *communication length* and is worst-case: we pay for the number of binary digits (*bits*) in the longest possible exchange (i.e. over all possible private data realizations).[4] Thus we are emphasizing that communication takes time and that this is a principal part of its cost.[5] A

1 Such an algorithm bears a formal resemblance to "management by exception" in which more communication occurs when private information takes on "exceptional" values. (Marschak and Radner [1972: 206–207])

2 This phenomenon can also occur in Dow's [1991] limited-memory search model. No communication from Past→Present is equivalent to buying the good from the first firm visited without bothering to visit the second firm to compare prices. Dow prohibits buying without observing both firms' prices. Kofman and Ratliff [1991b] show that the shopper would be strictly better off if she could waive the right to comparison shop. By purchasing without comparing prices when the first firm's price is very low (which is almost certainly the correct decision and is not a very costly mistake even when it is incorrect), she husbands her memory to more finely discriminate between higher prices for the first firm. This improves her decision about which firm to buy from in the situations in which the wrong decision would be more costly.

3 We thank Tom Marschak for posing to us a discrete formulation of this problem from which the present paper evolved.

4 A similar problem has been studied in the computer science literature concerning distributed and parallel processing. It takes the discrete form of approximating a Boolean-valued function defined on a grid. See Yao [79], Abelson [80], and Karp and Ng [forthcoming].

5 This information theoretic approach to measuring communication is used by Oniki [1986], who compares the informational efficiency of an auctioneer-mediated market with that of a centralized system for a particular single-good production economy. This contrasts with another strand in the literature in which the communication constraint is the dimensionality of the message space (i.e. the number of

mechanism is *efficient* if it results in a worst-case *decision error* no greater than that of any other mechanism which pays for the same number of bits.[1] For a given n we search for an efficient n-bit algorithm in order to establish the minimum worst-case decision error associated with n bits. By varying n we then can construct the *efficient error–communication length frontier*.

We now present two scenarios as exemplar problems which can be analyzed using the tools we develop in this paper.

Example 1: Team-spying the explosives factory

Two spies, Boris and Natasha, will be placed behind enemy lines to determine the amount of explosives being produced by a factory. The explosives are manufactured from two inputs X and Y according to a well-known production function. These inputs are produced in separate regions and hence arrive at the factory along different roads. The maximum capacity of each road to handle truck traffic is well known. From his vantage point Boris can observe with complete precision the flow of X into the factory but cannot observe at all the amount of Y; similarly Natasha can observe the flow of Y but not that of X. (See Figure 1.)

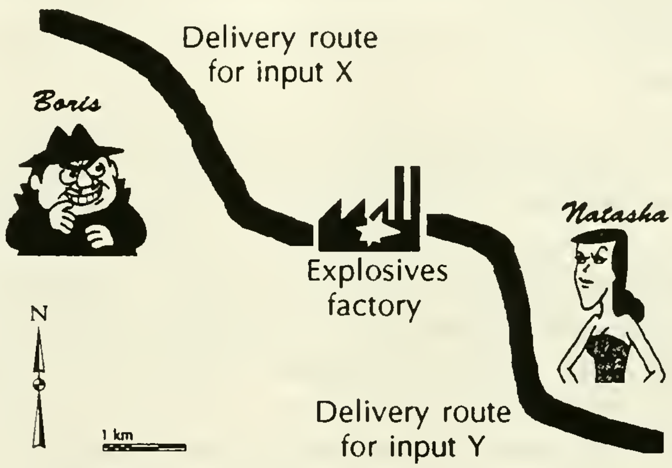

Figure 1: The map for the mission.

Each spy is equipped with a short-range radio which can send on/off pulses to the other. The enemy is not yet aware that the factory is being surveilled, and the more numerous are the inter-spy pulses, the greater is the risk that the espionage mission will be discovered. Consequently the spies want to minimize their communication with each other subject to a constraint on the accuracy of their production estimate. The radios' range is too short to reach their home territory, so one of the spies must physically

---

message "pipelines") with no constraint on how much a pipeline can be used. For example see Hurwicz [1960], Mount and Reiter [1974], Walker [1977], Osana [1978], Jordan [1982], and Green and Laffont [1987].

1 Minimizing the worst-case error is a reasonable optimality criterion in circumstances of "complete ignorance" (Arrow and Hurwicz [1977]) or in an infinite risk aversion limit. Its tractability aids in breaking new ground, and results achieved here generate conjectures to guide research under other criteria.

return home—by navigating the only available single-person kayak down a wild river—in order to deliver an intelligence assessment to their commanders.

Prior to being deployed in hostile territory Boris and Natasha meet at a local bistro to have a beer and to design a communication algorithm. Upon what rules of communication and decision delegation should they coordinate? For some fixed number n of pulses who should send the first pulse? Should it be followed by a second pulse from the same spy or perhaps by a premier pulse from the other? Who will be the whitewater courier delivering the final guestimate of explosives production? How will this sequence of pulses and the identity of the kayaker depend on the spies' private information (i.e. upon the quantities X and Y they each observe)?

Once they have decided upon a communication algorithm for every fixed number n of pulses and determined its worst-case error, there is still the question of what number n to pick. More pulses would result in a greater chance that their cover will be blown but would also increase the accuracy of their intelligence report. What is the tradeoff? I.e. how much would accuracy be improved by an increase in the inter-spy communication?

Example 2: Production decisions in an informationally decentralized firm

Consider a ski equipment manufacturer where the personnel director must be told how many workers w to hire for the upcoming production season. The relevant information for the optimal decision is distributed throughout the organization: The production manager will learn precisely the productivity-per-worker $\alpha \in [0, \bar{\alpha}]$. The marketing chief will learn the precise value of a demand parameter $\beta \in [0, \bar{\beta}]$. (See Figure 2.)

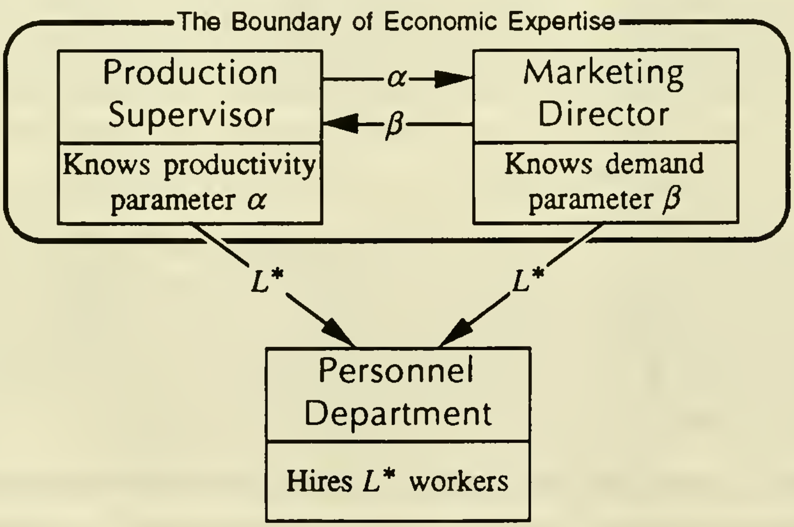

Figure 2: An informationally decentralized firm.

When communication is costly, the design problem is to construct an efficient procedure by which the production manager and marketing chief will share their private information in order to estimate w^*.[1] After this exchange the appropriate party can inform the personnel director of the optimal employment level w^*.[2] The goal is to minimize the worst-case departure from optimal employment over all possible productivity and demand combinations.

The same design questions arise here as in the spy example. For a fixed communication cost, who should speak first, the production manager or the marketing chief? Should they discuss the matter in a dialog? Or should the production manager, say, use the entire communication allotment to inform her partner as precisely as possible about productivity-per-worker so that the marketing chief can then calculate an approximation to the optimal employment level w^*? After this design problem has been solved for an arbitrary fixed communication cost, the decision remains about how much communication to perform. What is the tradeoff between communication cost and worst-case employment error? At what point does it no longer pay for the two informed parties to continue to talk to one another?

On interpretation

The phenomena with which we are concerned are communication, coordination, and decision delegation. Of course we realize that people do not communicate by exchanging bits in order to locate each other one-dimensionally on an interval, that the problem of an organization is not to announce a value which minimizes the worst-case error of approximation to a function, and that incentive compatibility is a crucial constraint. We do, however, see our model as a usefully interesting, yet tractable, metaphor for more complex situations.

The more general scenario we have in mind is two people seeking to arrive at a consensus action which will bring them satisfaction if the action is appropriate to their common situation and some pain if it is not. The focus of our research is the process by which two individual and privately known situations become a more nearly commonly known situation. That is to say: how do two agents come to share an understanding of the state of the world?

We believe that there is no way in which a human being can express to another the full meaning of her situation. We claim that a person has no method of letting someone else know everything that she perceives, believes, remembers, guesses, understands, fears, hopes, and doubts about both the environment and herself in a finite amount of time. However, action must be taken after only a finite conversation, so we must be concerned with efficient communication.

For example, the production manager does not only know the normal productivity of different inputs for a given technology. She also knows the reliability and morale of the workers, the probability that the

---

1 The personnel director rarely attended his economics classes and would not know how to compute the optimal employment level even if he were perfectly informed about α and β; thus he is outside the boundary of economic expertise. Consequently we need not consider algorithms in which the production manager and the marketing chief both communicate approximations of their private information directly to the personnel director.

2 This communication with the personnel director is also costly. However, we assume that it is equally costly regardless of whether it is done by the production manager or the marketing chief. Therefore it is a fixed cost in the design problem.

machines will break down, the necessary maintenance schedule, the gossip about new technical developments, etc. (Moreover, she has a lot of background beliefs of which she may not even be aware!) We assert that she cannot transmit all this, so she must use her language providently when she coordinates her actions with others.

In our model we do not attempt to replicate a complicated, multidimensional "common situation;" we consider points in a rectangle. On the other hand, we need not endow the agents with a rich context-dependent natural language; we have zeros and ones. In our opinion tractability justifies the simplicity we impose when we project what-to-communicate and how-to-communicate onto a conceptually manageable space. We can then address issues of syntax (the rules of communication) and semantics (the meaning of the messages) within an elegant mathematical formalism.[1]

2. The Model

A team has two partners: X and Y. They have a joint address $\theta \equiv (\theta_X, \theta_Y) \equiv (x, y)$ on a rectangle

$$E = E_X \times E_Y, \tag{A.1}$$

where E_X and E_Y are the closed intervals

$$E_X \equiv [a, b], \tag{A.2a}$$

$$E_Y \equiv [c, d]. \tag{A.2b}$$

A real-valued *optimal action function* is defined and bounded on the rectangle E, viz. $\varphi\colon E \to \mathbb{R}$, and is known to both players. Ideally the team would take the optimal action $\varphi(\theta)$ corresponding to their joint address. However, each partner knows only a projection of θ; X knows the precise value of x and Y knows the precise value of y. Initially X's only information about Y's location is that $y \in E_Y$. Similarly, Y knows only that $x \in E_X$.

The partners may communicate by sending binary digits (*bits*) to one another. One of the players then chooses an action $\psi \in \mathbb{R}$ which approximates the desired action $\varphi(\theta)$. The communication process and the choice of action is dictated by a *communication algorithm* R, which is understood by both players. The algorithm R must specify the *instigator*—the partner who initiates the process by either immediately selecting an action or by sending the first bit—as well as who sends what bits to whom and when and who chooses the ultimate action and when.[2] Because knowledge of x and y is privately held by X and Y, respectively, partner i's activities under the algorithm R can depend only on her own coordinate θ_i and on the history of bits sent and received; they cannot depend explicitly on her partner's coordinate θ_j. Without loss of generality we assume that the two partners never send simultaneous messages.

---

1 Our strategy of radical simplification seems to be vindicated by our work-in-progress using a multidimensional information structure for each partner. We have found interesting multidimensional problems which decompose into separate one-dimensional problems and which therefore directly require the techniques we develop in this paper.

2 Because of the information partition the identity of the instigator must be a property of the algorithm which is independent of θ. (Otherwise there would be cases in which both players tried to instigate or neither instigated.)

For a given θ the communication algorithm R prescribes a particular exchange of some number $\eta(\theta)$ of bits and results in some action $\psi(\theta)$. The *communication length*, n, of R is the number of bits sent in the longest possible exchange, i.e.

$$n = \max_{\theta \in E} \eta(\theta). \tag{A.3}$$

Note that we are adopting a *worst-case* definition of the communication length of the algorithm R. The *error of the algorithm R* is also a worst-case formulation:

$$\varepsilon = \sup_{\theta \in E} |\psi(\theta) - \varphi(\theta)|. \tag{A.4}$$

Without loss of generality we assume that ψ is bounded on the unit square to ensure that the error as a function of θ is bounded and therefore that the error of R is defined.

The designer's problem is to specify—in ignorance of which θ will be realized—communication algorithms which are *efficient* with respect to communication length and error of approximation. An algorithm R is efficient if there exists no other algorithm which achieves either a strictly lower error with a weakly lower communication length or a strictly lower communication length with a weakly lower error. We wish to characterize for a given function φ the efficient frontier of achievable (error, communication length) pairs, i.e. those which result from efficient algorithms. We need not find all of the efficient algorithms; we need locate only a subset sufficient to demarcate the frontier.

For a given joint location θ, the algorithm R will ultimately call upon one of the partners to be the *actor for* θ. We characterize the actor's knowledge about θ at that time by specifying the actor's *information set*: the minimal subset $S \subset E$ within which she knows θ to lie. Given her knowledge that $\theta \in S$, the actor knows that the optimal action $\varphi(\theta)$ belongs to the image of S under φ:

$$\varphi(S) \equiv \{\varphi(\theta) : \theta \in S\}. \tag{A.5}$$

The actor's best (i.e. worst-case error minimizing) approximation action given S is the midpoint of $\varphi(S)$,

$$\psi(S) = \tfrac{1}{2}(\inf \varphi(S) + \sup \varphi(S)), \tag{A.6}$$

which results in the worst-case error given S,

$$\varepsilon(S) = \begin{cases} \tfrac{1}{2}\Delta\varphi(S), & S \neq \varnothing, \\ 0, & S = \varnothing, \end{cases} \tag{A.7}$$

where we denote the function's *oscillation* over S by

$$\Delta\varphi(S) \equiv \sup \varphi(S) - \inf \varphi(S). \tag{A.8}$$

We observe that adding points to S would weakly increase the supremum and weakly decrease the infimum and thereby weakly increase the oscillation of φ over S; therefore the error over a strictly larger set is weakly greater than the error over the subset.

In section 3 we discuss monolog algorithms—those which dictate that all communication is performed by the instigator. We restrict attention to optimal action functions which satisfy appropriate monotonicity and continuity assumptions. For a given communication length we compute in "almost closed form" the error and structure of algorithms which are efficient within the class of monologs. We begin section 4 by presenting two examples which demonstrate that monolog need not be sufficient for efficiency—a dialog, in which both partners speak, can outperform a best monolog. We then establish a sufficient condition for monolog efficiency: additive separability of the optimal action function. We show that this condition is not necessary by exhibiting a rectangle on which a multiplicatively separable optimal action function can be efficiently approximated through the use of a monolog. In section 5 we show that all of our results are valid even when the monotonicity requirements are dropped as long as the optimal action function is appropriately separable. Section 6 summarizes the paper and gathers our final thoughts.

3. Monolog algorithms

We will find it useful to study a simple subset of the set of communication algorithms: the set of *monolog* communication algorithms. Such an algorithm stipulates that for all θ any transmission of bits is performed only by the instigator. If partner i, say, is the instigator, she either immediately chooses an action $\psi(\theta_i)$ or chooses from among 2^n messages. (In a monolog any bits sent are transmitted as an uninterrupted string.) If the instigator chooses message κ, then her partner, j, chooses an action $\psi(\theta_j, \kappa)$ which depends on his coordinate and the message he received.

For definiteness we consider here monologs in which X is the instigator. A parallel analysis applies to Y-instigator monolog algorithms. In order to find an algorithm which is efficient within the class of monologs we would find a monolog which is efficient within the class of X-instigator monologs and a monolog which is efficient within the class of Y-instigator monologs. An algorithm with the lower error would be efficient within the broader class of monolog algorithms. For the remainder of our discussion of X-instigator monologs we will use the term "efficient" as a shorthand to mean "efficient within the class of X-instigator monolog algorithms."

We first describe a monolog as a partition of E_X and discuss the role of immediate action. We give conditions under which we can restrict attention to partitions whose cells are convex. We discuss the efficiency and existence of partitions which, loosely speaking, yield a common error in every cell. We then solve the efficient monolog problem—computing both the error and the algorithm itself—for a large class of optimal action functions and discuss the comparative statics of increasing the communication budget.

Monolog algorithms as partitions

We can fully define a monolog by a (2^n+1)-cell partition of E_X—which specifies the circumstances under which the instigator should immediately act or send a particular message—and by choice-of-action rules which specify how each partner would choose an action if called upon to do so. We define

$$m = 2^n, \tag{B.1}$$

$$M = \{1, \ldots, m\}, \tag{B.2}$$

$$M^\dagger = M \cup \{0\}. \tag{B.3}$$

The $(m+1)$-cell partition of E_X is denoted

$$\sigma = \{\sigma_0, \sigma_1, \ldots, \sigma_m\}, \tag{B.4}$$

i.e. where each $\sigma_i \subset E_X$, $\cup_{i \in M^\dagger} \sigma_i = E_X$, and $\sigma_i \cap \sigma_j = \varnothing$ when $i \neq j$.

The cell σ_0 is the *immediate action cell.* When $x \in \sigma_0$, X is called upon to immediately take an action. She knows only that $\theta \in \{x\} \times E_Y$ and therefore that the optimal action lies in the image $\varphi(\{x\} \times E_Y)$. She takes the action $\psi(\{x\} \times E_Y)$ specified in (A.6), which results in the error $\varepsilon(\{x\} \times E_Y)$ from (A.7). (See Figure 3.) The error ε_0 for the immediate action cell is

$$\varepsilon_0 = \begin{cases} \sup\limits_{x \in \sigma_0} \varepsilon(\{x\} \times E_Y), & \sigma_0 \neq \varnothing, \\ 0, & \sigma_0 = \varnothing. \end{cases} \tag{B.5}$$

Adding points to σ_0 could only increase the supremum in (B.5); therefore the error of immediate action would be weakly greater for a strictly larger immediate action cell.

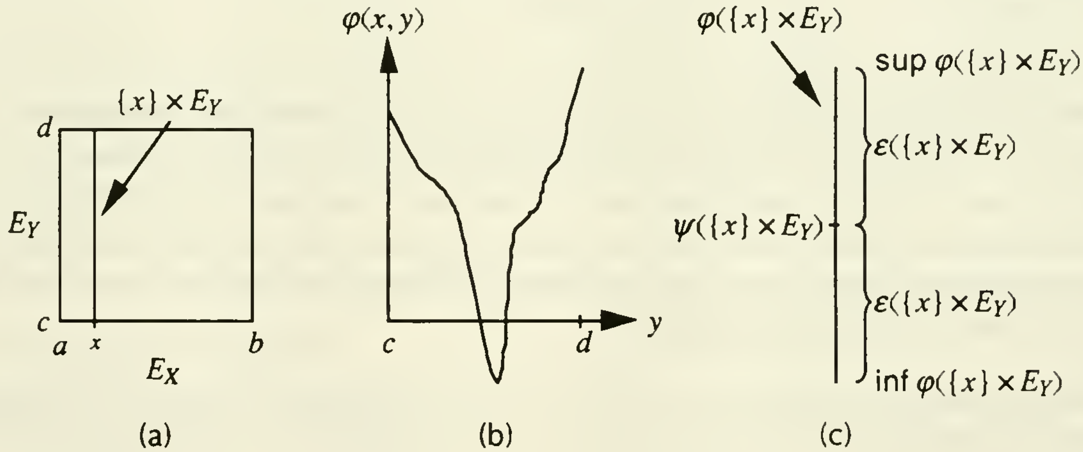

Figure 3: (a) The set of θ, (b) the image under φ, and (c) the best-approximation and the worst-case error when X immediately announces at x.

The remaining m cells are the *message cells*. When $x \in \sigma_i$, $i \in M$, X sends the message i in order to inform Y that $x \in \sigma_i$. Y then knows that $\theta \in \sigma_i \times \{y\}$ and therefore takes the action $\psi(\sigma_i \times \{y\})$, which results in the error $\varepsilon(\sigma_i \times \{y\})$. (See Figure 4.) The worst-case error ε_i when X sends message i is

$$\varepsilon_i = \sup_{y \in E_Y} \varepsilon(\sigma_i \times \{y\}). \tag{B.6}$$

If we were to add points to σ_i, then for each y the error $\varepsilon(\sigma_i \times \{y\})$ would weakly increase, causing the supremum of these errors to weakly increase as well; therefore the error of a strictly larger message cell is weakly greater than the error of a smaller message cell. If there exists a y^* at which the worst-case error is achieved, i.e. such that

$$y^* \in \arg\max_{y \in E_Y} \varepsilon(\sigma_i \times \{y\}), \tag{B.7}$$

we say that y^* is a *worst-case y value for* σ_i.

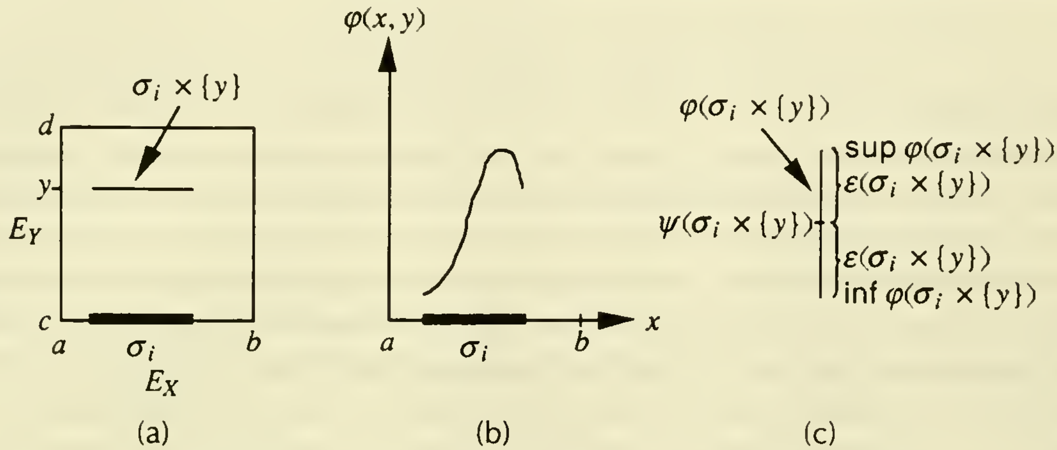

Figure 4: (a) The set of θ, (b) the image under φ, and (c) the best-approximation and the worst-case error when X sends message i for a given y.

The error of the algorithm σ is the largest individual cell error, i.e.

$$\varepsilon = \max\{\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_m\}. \tag{B.8}$$

The designer's task is to choose the partition σ of E_X which minimizes ε.

Immediate action and efficiency

If the immediate action cell σ_0 of an algorithm partition σ is empty, we say that σ is a $1 \to m$ partition; if $\sigma_0 \neq \varnothing$, we say that σ is a $0 \to m$ partition. When does efficiency require an empty immediate action cell? To answer this question we now turn to a significant difference between an immediate action cell and a message cell: If a message cell σ_i, $i \in M$, is a singleton, i.e. consists of a single point, its associated error ε_i is zero because, after receiving message i, Y would know precisely not only his own coordinate y but X's coordinate x as well. Therefore if some message cell is empty, it would not increase the error of the algorithm if we were to transfer a single point to the empty message cell from another cell. Therefore efficiency can never require that some message cell be empty. On the other hand if the immediate action cell σ_0 is a singleton, say $\sigma_0 = \{x\}$, the associated error $\varepsilon(\{x\} \times E_Y)$ can be positive because X has uncertainty about Y's coordinate.

Consider a $1 \to m$ partition σ whose error is ε. If for some $\hat{x} \in E_X$ the error of immediate action is less than the error of the $1 \to m$ partition, i.e. if $\varepsilon(\{\hat{x}\} \times E_Y) < \varepsilon$, then it would weakly decrease the algorithm's error if we removed $\hat{x}$ from its message cell and added it to the previously empty immediate action cell. We would say the $1 \rightarrow m$ partition is *improvident* because it squanders its opportunity to weakly reduce its error through the incorporation of a nonempty immediate action cell. If on the other hand for all $x \in E_X$ the immediate action error exceeds the error of the $1 \rightarrow m$ partition, then it would strictly increase the algorithm's error to incorporate even a single point into its empty immediate action cell. In this case we say that the $1 \rightarrow m$ partition is *provident*; eschewing immediate action is efficient.

Convex algorithm partitions

We now consider *convex algorithm partitions*—those of the form $\sigma = \{\sigma_0, \sigma_1, \ldots, \sigma_m\}$ such that each σ_i, $i \in M^\dagger$, is a convex set (i.e. an interval). Convex partitions are especially tractable compared to more general partitions of $[a, b]$ because they can be almost completely specified by a set of m numbers representing the cells, e.g. $\sigma_0 = [a, x_0]$, $\sigma_i = (x_{i-1}, x_i]$, $i \in M$, $x_m = b$.[1] In order to take advantage of their structural simplicity we will in Theorem 1 impose monotonicity conditions upon the optimal action function φ which guarantee that we can without loss of generality restrict attention to convex partitions; i.e. if presented with a partition which is not convex, we can exhibit a convex partition whose error is weakly less than that of the original partition. (These monotonicity assumptions, as well as additional assumptions imposed later, are not as restrictive as they might seem. We will see in Theorem 8 that all of our results can be easily extended to a much wider class of optimal action functions.)

Conditions under which convex partitions are sufficient

Theorem 1 For each $y \in E_Y$, let $\varphi(x, y)$ and $\varepsilon(\{x\} \times E_Y)$ be weakly monotonic functions of x on E_X, and let $\sigma = \{\sigma_0, \sigma_1, \ldots, \sigma_m\}$ be a partition of E_X whose error [from (B.8)] is ε. Then there exists a convex partition $\hat{\sigma} = \{\hat{\sigma}_0, \hat{\sigma}_1, \ldots, \hat{\sigma}_m\}$ of E_X, whose error we denote by $\hat{\varepsilon}$, such that ① The error of $\hat{\sigma}$ is no greater, i.e. $\hat{\varepsilon} \leq \varepsilon$; and ② if the immediate action interval $\hat{\sigma}_0$ is nonempty, it will be at the left (respectively right) end of E_X if $\varepsilon(\{x\} \times E_Y)$ is weakly increasing (respectively weakly decreasing), i.e. $\hat{\sigma}_0 \ni a$ (respectively $\hat{\sigma}_0 \ni b$).

Proof For simplicity of exposition assume that φ and $\varepsilon(\{\cdot\} \times E_Y)$ are continuous and that the monotonicity of $\varphi(\cdot, y)$ has the same sense for all $y \in E_Y$.[2] For definiteness assume that both of the monotonicity assumptions are satisfied in the weakly increasing sense. First we demonstrate claim ②. If an algorithm specifies an immediate action for some x_0, the monotonicity of $\varepsilon(\{x\} \times E_Y)$ implies that the error of immediate action for any $x < x_0$ would be weakly lower; therefore the error of the algorithm would not be increased by incorporating within the immediate action cell all values in the interval $[a, x_0]$. Therefore we can replace σ_0 by $[a, \sup \sigma_0]$ without increasing the error of immediate action.

Now we show ①. Consider the i-th message cell. Thanks to monotonicity (and temporarily assumed continuity) we have

---

1 These m numbers do not themselves specify the particular closedness decisions for each cell. When we make an additional continuity assumption, the significance of the closedness decisions is nil.

2 More general and more detailed versions of the proofs offered here appear in Kofman and Ratliff [1991a].

$$\sup \varphi(\sigma_i \times \{y\}) = \varphi(\sup \sigma_i, y),$$

$$\inf \varphi(\sigma_i \times \{y\}) = \varphi(\inf \sigma_i, y).$$

Therefore for any $y \in E_Y$, the error when Y acts after receiving message i is

$$\varepsilon(\sigma_i \times \{y\}) = \tfrac{1}{2}(\varphi(\sup \sigma_i, y) - \varphi(\inf \sigma_i, y)),$$

whose only dependence upon σ_i is through its supremum and infimum, which would be unchanged if we convexified this set. Because $\varepsilon(\sigma_i \times \{y\})$ would be unchanged for every y by convexification of σ_i, the error ε_i from (B.6) would be unchanged as well. We construct $\hat{\sigma}$ by refining the covering $\{\operatorname{co}\sigma_0, \operatorname{co}\sigma_1, \ldots, \operatorname{co}\sigma_m\}$ in such a way that each $\hat{\sigma}_i$ is convex and $\hat{\sigma}_i \subset \operatorname{co}\sigma_i$. ☺

The importance of monotonicity is exemplified in Figure 5 in which two optimal action functions φ^{mono} and φ^{non}, horizontally monotonic and nonmonotonic, respectively, are plotted as functions of x for an arbitrarily chosen y. The nonconvex subset σ_i is the union of two disjoint intervals. In the lower half of the figure—the horizontally monotonic case—it is obvious that convexifying σ_i does not increase the oscillation of φ^{mono} and therefore does not increase the error, i.e. $\Delta\varphi(\sigma_i \times \{y\}) = \Delta\varphi(\operatorname{co}\sigma_i \times \{y\})$ and therefore $\varepsilon(\sigma_i \times \{y\}) = \varepsilon(\operatorname{co}\sigma_i \times \{y\})$. In the upper half of Figure 5 the image of the nonmonotonic φ^{non} over σ_i is a single point and hence the oscillation and error are zero. After convexification the image is the indicated segment of values, and therefore the oscillation and error are positive; i.e. the inclusion of new points through convexification results in an increased error over the new domain, i.e. $\varepsilon(\operatorname{co}\sigma_i \times \{y\}) > \varepsilon(\sigma_i \times \{y\})$.

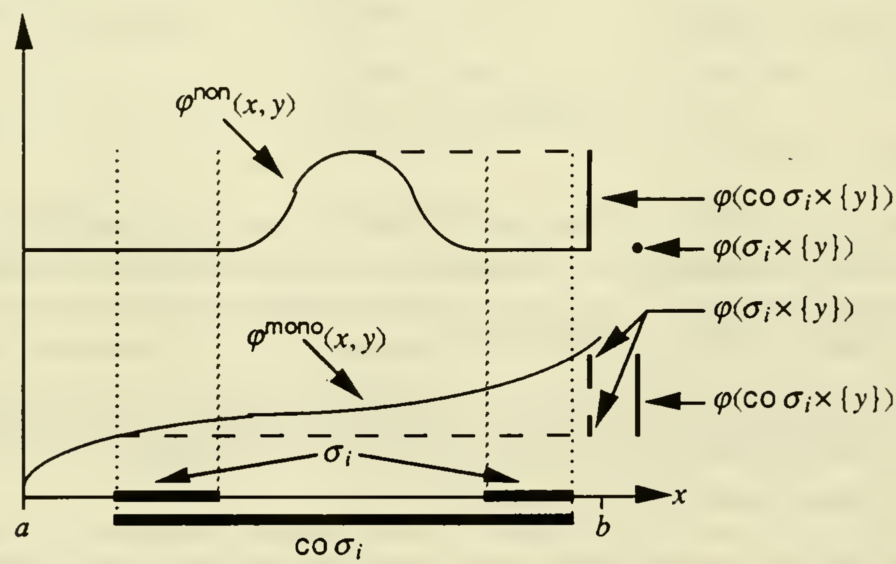

Figure 5: The worst-case error for a horizontally monotonic function is unaffected by convexification; this need not be true in the absence of monotonicity.

Sufficient conditions for the monotonicity of the immediate action error

The requirement that φ be weakly horizontally monotonic is easily interpreted; the requirement that $\varepsilon(\{x\} \times E_Y)$ be a weakly monotonic function of x (which is indirectly a restriction upon φ) does not have as obvious an interpretation. The next theorem states sufficient conditions when φ is appropriately differentiable which guarantee the monotonicity of the immediate action error.[1]

Theorem 2 Let $\varphi(\cdot, y)$ be either ① a weakly increasing function for all $y \in E_Y$ or ② a weakly decreasing function for all $y \in E_Y$. Further let $\varphi(x, \cdot)$ be either ① weakly increasing for all $x \in E_X$ or ② weakly decreasing for all $x \in E_X$. Let $\partial\varphi/\partial x$, $\partial\varphi/\partial y$, and $\partial^2\varphi/\partial x\partial y$ exist and be continuous over E. Then the following two statements are equivalent:

① $\partial^2\varphi/\partial x\partial y$ has weakly constant sign over E (i.e. nonnegative over E or nonpositive over E).

② The errors $\varepsilon(\{x\} \times \sigma_Y)$ and $\varepsilon(\sigma_X \times \{y\})$ are weakly monotonic functions of x and y, respectively, for every choice of subsets $\sigma_X \subset E_X$ and $\sigma_Y \subset E_Y$.

Specifically, when the mixed partial is {nonnegative | nonpositive}, X's error $\varepsilon(\{x\} \times \sigma_Y)$ is weakly {increasing | decreasing} as φ is vertically weakly {increasing | decreasing} and Y's error $\varepsilon(\sigma_X \times \{y\})$ is weakly {increasing | decreasing} as φ is horizontally weakly {increasing | decreasing}.

Proof We will first show that ① implies the first claim of ② when $\varphi_{xy} \geq 0$ and $\varphi_y \geq 0$. The other cases are demonstrated similarly. Let $\inf \sigma_Y = c$ and $\sup \sigma_Y = d$. Because $\varphi_y \geq 0$, $\varepsilon(\{x\} \times \sigma_Y) = \frac{1}{2}(\varphi(x, d) - \varphi(x, c))$. This is weakly increasing in x because $\varphi_x(x, d) \geq \varphi_x(x, c)$ for all $x \in E_X$ thanks to $\varphi_{xy} \geq 0$. (The continuity of φ_{xy} is used to guarantee that $\varphi_{xy} = \varphi_{yx}$.)

We now show that ② ⇒ ① for the case in which $\varphi_y \geq 0$ and in which $\varepsilon(\{x\} \times \sigma_Y)$ is weakly increasing. Because $\varepsilon(\{x\} \times \sigma_Y)$ is weakly increasing, we know that $\varphi_x(x, d) \geq \varphi_x(x, c)$ for all $x \in E_X$ and all $\sigma_Y \subset E_Y$. Therefore

$$\varphi_{xy}(x, y) \equiv \lim_{h \to 0} \frac{\varphi_x(x, y+h) - \varphi_x(x, y)}{h} \geq 0,$$

for all $x \in E_X$ and $y \in \operatorname{int} E_Y$. The result extends to all of E due to the continuity of φ_{xy}. ☺

Equal-error convex partitions

For the remainder of our analysis of monolog algorithms we restrict attention to optimal action functions φ which satisfy the hypotheses of Theorem 1. Therefore we can without loss of generality confine our consideration to convex partitions.

We now direct our analysis toward those convex partitions which result, loosely speaking, in the same worst-case error in every cell. Theorem 3 establishes conditions under which we can be certain that any given "equal-error" convex partition is efficient. Theorem 4 establishes conditions under which we can be certain that such an equal-error convex partition exists.

---

[1] The implications of these sufficient conditions are obviously stronger than we need to apply Theorem 1 to X-instigator algorithms; they are applicable when Y instigates as well. In addition, these conditions will be sufficient for the satisfaction of stronger hypotheses of later theorems.

We say that σ is an *equal-error* $0 \to m$ partition if

$$\varepsilon_0 = \varepsilon_1 = \cdots = \varepsilon_m \qquad \text{and} \quad \sigma_0 \neq \varnothing. \tag{B.9}$$

We say that σ is an *equal-error* $1 \to m$ partition if

$$\varepsilon_1 = \cdots = \varepsilon_m \qquad \text{and} \quad \sigma_0 = \varnothing. \tag{B.10}$$

We say that σ is an *equal-error partition* if it is either an equal-error $0 \to m$ partition or an equal-error $1 \to m$ partition. We have previously defined the terms provident and improvident with respect to $1 \to m$ partitions. We now define any equal-error $0 \to m$ partition to be provident.

If a convex, provident, equal-error partition exists, it is efficient

Theorem 3 will tell us that if a convex, provident, equal-error partition exists then it must be efficient; i.e. no other partition with the same communication length could possibly have a lower error. As an example, consider a convex partition with two message cells, $\sigma = \{\sigma_L, \sigma_R\}$, where $\sigma_L = [a, x_1)$ and $\sigma_R = [x_1, b]$, which we assume to be equal-error with $\varepsilon = \varepsilon_L = \varepsilon_R$. Assume further that, for all $x \in E_X$, $\varepsilon(\{x\} \times E_Y)$ is sufficiently large that this partition is provident (i.e. efficiency does not require a nonempty immediate action cell). Now consider any other convex partition with two message cells, $\tilde{\sigma} = \{\tilde{\sigma}_L, \tilde{\sigma}_R\}$, with error $\tilde{\varepsilon} = \max\{\tilde{\varepsilon}_L, \tilde{\varepsilon}_R\}$. Let its intercell boundary be $\tilde{x}_1$. This new partition cannot have a lower error. If it did we would have both $\tilde{\varepsilon}_L < \varepsilon$ and $\tilde{\varepsilon}_R < \varepsilon$. In order to make $\tilde{\varepsilon}_L < \varepsilon$, we would have to choose $\tilde{x}_1 < x_1$ in order to shrink $\tilde{\sigma}_L$ relative to σ_L. However, thus shrinking $\tilde{\sigma}_L$ makes $\tilde{\sigma}_R$ a larger set than σ_R and therefore $\tilde{\varepsilon}_R \geq \varepsilon$.

Theorem 3 For each $y \in E_Y$, let $\varphi(x, y)$ and $\varepsilon(\{x\} \times E_Y)$ be weakly monotonic functions of x on E_X. Let σ be a provident, equal-error, convex partition. Then there does not exist a partition (whether $0 \to m$ or $1 \to m$) which has a lower error.

Sketch of Proof The proof of Theorem 3, given in full in Kofman and Ratliff [1991a], relies upon a lemma which states that if σ and $\tilde{\sigma}$ are two distinct partitions with the same number of cells, then some cell $\tilde{\sigma}_i$ of the second partition is a strictly larger set—and has a weakly larger error—than some cell σ_k of the first partition.[1] (If σ is a provident $1 \to m$ partition, then any potentially lower-error $\tilde{\sigma}$ would also have to be a $1 \to m$ partition because otherwise $\tilde{\varepsilon} \geq \tilde{\varepsilon}_0 \geq \inf_{x \in [a,b]} \varepsilon(\{x\} \times E_Y) > \varepsilon$.) Therefore the error of this different partition will be at least as large as the common error of the equal-error partition. ☺

---

[1] Actually, a stronger claim is shown: that either the left-hand cell of $\tilde{\sigma}$ is strictly larger than the left-hand cell of σ, the right-hand cell of $\tilde{\sigma}$ is strictly larger than the right-hand cell of σ, or one of the interior cells of $\tilde{\sigma}$ is strictly larger than one of the interior cells of σ. This assures us that we are not comparing errors based on a subset relationship between a message cell and an immediate action cell.

Continuity implies existence of an equal-error partition

By requiring continuity of the optimal action function we are guaranteed that an equal-error partition exists.

Theorem 4 Let φ be continuous and for every $y \in E_Y$, let $\varphi(x,y)$ and $\varepsilon(\{x\} \times E_Y)$ be strictly monotonic functions of x. Then for all positive integers m, there exists a unique equal-error $1 \to m$ partition. If this $1 \to m$ partition is improvident, then there also exists a unique equal-error $0 \to m$ partition. (Therefore there always exists a provident equal-error partition.) The error of this $0 \to m$ partition is weakly less than the error of the $1 \to m$ partition.

Sketch of Proof This theorem probably seems intuitively plausible to the reader. The full proof, which is given in Kofman and Ratliff [1991a], has its conceptual substrate in the following iterative procedure, where for definiteness we assume that the $1 \to m$ partition is provident (the modification for the $0 \to m$ case is straightforward): Fix some sufficiently small ε and find the endpoint x_1 of $\sigma_1 = [0, x_1)$ such that $\varepsilon_1 = \varepsilon$ [see (B.6)]. Now find the endpoint x_2 of $\sigma_2 = [x_1, x_2)$ such that $\varepsilon_2 = \varepsilon$. Continue in this fashion until either a $x_m = 1$ is determined, b $x_m < 1$ is determined, or c the right-hand boundary of $[a, b]$ is reached and there are cells σ_i, $i \le m$, remaining which have yet to receive their ε error allotment. In case a, the desired unique equal-error $1 \to m$ partition has been found. In cases b and c, ε must be increased or decreased, respectively, and then the process repeated. By choosing ε sufficiently small or large, respectively, we can achieve cases b and c. The continuity hypotheses guarantee that $x_m(\varepsilon)$ is a continuous function of ε and therefore that a value for ε can be found which will result in the desired case a. ☺

Theorem 4 is an existence result. However, it is easy to see how the iterative procedure described in the above sketch of the proof provides an algorithm to actually compute an efficient monolog algorithm given an optimal action function satisfying the assumptions of the theorem.

The theorem suggests the following strategy to find an efficient algorithm: First find an equal-error $1 \to m$ partition and determine its error $\check{\varepsilon}$. To determine whether this partition is provident find the minimum error of immediate action. If this error is weakly larger than $\check{\varepsilon}$, the $1 \to m$ partition is efficient. Otherwise, find an equal-error $0 \to m$ partition; it will be efficient.

Monolog error under horizontal additivity

The task of finding an efficient monolog is greatly eased when the optimal action function satisfies an additional vertical monotonicity assumption. (Satisfaction of this assumption is implied by the constant-signed mixed-partial requirement of Theorem 2.) We can then determine the error over any message cell σ_i, $i \in M$, simply by evaluating the optimal action function φ along a single, worst-case horizontal segment. This simplification yields "almost closed form" expressions for the error and the cell boundaries of an efficient partition.

Theorem 5 Let φ be continuous and let $\varphi(\cdot, y)$ be either weakly increasing for all $y \in E_Y$ or weakly decreasing for all $y \in E_Y$. Further, let φ be such that the error, given y, when X sends message i, viz. $\varepsilon(\sigma_i \times \{y\})$, is either ⓐ weakly increasing in y for all $\sigma_i \subset E_X$ or ⓑ weakly decreasing in y for all $\sigma_i \subset E_X$. Define the *worst-case y-value* $y^* = d$ in case ⓐ and $y^* = c$ in case ⓑ.

① There exists an equal-error $1 \to m$ partition σ of $[a, b]$ whose error is

$$\check{\varepsilon} = \frac{|\varphi(b, y^*) - \varphi(a, y^*)|}{2m}. \tag{5.1}$$

Define $\tilde{\varphi}(x) \equiv \varphi(x, y^*)$. The cell boundaries of σ are determined (uniquely if the horizontal monotonicity of φ is strict) by

$$\sup \sigma_k \in \tilde{\varphi}^{-1}(\tilde{\varphi}(\min \sigma_1) \pm 2k\check{\varepsilon}), \tag{5.2}$$

$k \in M$, and by $\min \sigma_1 = a$, where the sense of the "±" corresponds to $\varphi(\cdot, y^*)$ increasing and decreasing, respectively.[1]

② Let $\varepsilon(\{x\} \times E_Y)$ be a weakly monotonic function of x. Define the best immediate announcement point $\hat{x}_{\text{best}} = a$ (respectively $\hat{x}_{\text{best}} = b$) if $\varepsilon(\{x\} \times E_Y)$ is weakly increasing (respectively weakly decreasing). Define the immediate action error function to be

$$\varepsilon_0(x_0) = \frac{|\varphi(x_0, d) - \varphi(x_0, c)|}{2}. \tag{5.3}$$

If $\varepsilon_0(\hat{x}_{\text{best}}) \geq \check{\varepsilon}$, then σ is efficient. Otherwise, there exists an equal-error $0 \to m$ partition $\hat{\sigma}$ of $[a, b]$. Define $\hat{x}_m = b$ (respectively $\hat{x}_m = a$) when $\hat{x}_{\text{best}} = a$ (respectively $\hat{x}_{\text{best}} = b$) and define the message cell error function to be

$$\varepsilon(x_0) = \frac{|\varphi(\hat{x}_m, y^*) - \varphi(x_0, y^*)|}{2m}. \tag{5.4}$$

There exists $\hat{x}_0 \in E_X$ such that $\varepsilon_0(\hat{x}_0) = \varepsilon(\hat{x}_0)$, and the error of $\hat{\sigma}$ is $\hat{\varepsilon} \equiv \varepsilon_0(\hat{x}_0) = \varepsilon(\hat{x}_0) \leq \check{\varepsilon}$. The immediate action interval is $\hat{\sigma}_0 = \text{co}\{\hat{x}_{\text{best}}, \hat{x}_0\}$. The cell boundaries of an efficient partition are again determined by (5.2)—where $\check{\varepsilon}$ is replaced by $\hat{\varepsilon}$—and by $\min \sigma_1 = \hat{x}_0$ (respectively $\min \sigma_1 = a$) when $\varepsilon(\{x\} \times E_Y)$ is weakly increasing (respectively weakly decreasing).

Proof From the definition of y^* and the monotonicity of $\varphi(\cdot, y)$ and $\varepsilon(\sigma_i \times \{\cdot\})$, we see from (B.6) that the error over message cell i is

$$\varepsilon_i = \sup_{y \in E_Y} \varepsilon(\sigma_i \times \{y\}) = \varepsilon(\sigma_i \times \{y^*\}) = \frac{1}{2}|\varphi(\sup \sigma_i, y^*) - \varphi(\inf \sigma_i, y^*)|.$$

The partition is equal-error, so the common message-cell error is

[1] We use ϕ^{-1} to denote the inverse image of ϕ, viz. $\phi^{-1}(z) = \{x \in E_X: \phi(x) = z\}$.

$$\varepsilon=\frac{1}{m}\sum_{i\in M}\varepsilon_i=\frac{1}{2m}\sum_{i\in M}\left|\varphi(\sup\sigma_i,y^*)-\varphi(\inf\sigma_i,y^*)\right|=\frac{1}{2m}\left|\varphi(\sup\sigma_m,y^*)-\varphi(\inf\sigma_1,y^*)\right|,$$

where the message cells are labeled such that, for all $i,j\in M$, $i<j\Leftrightarrow\inf\sigma_i\le\inf\sigma_j$. This establishes (5.1) and (5.4). The error over the immediate action cell from (B.5) is given by (5.3). The equal-error requirement then determines $\hat{x}_0$. The boundaries defined by (5.2) are easily verified to result in equal-error cells. ☺

Example 3: A multiplicatively separable optimal action function

As an example we will apply Theorem 5 to the case where $\varphi(x,y)=xy$ on $E=[a,b]\times[c,d]\subset\mathbb{R}_+^2$. Let $\Delta x\equiv b-a$ and $\Delta y\equiv d-c$. We observe from Theorem 2 (because $\partial^2\varphi/\partial x\partial y\equiv 1$) that this function satisfies the hypotheses of the Theorem 5, including that $\varepsilon(\sigma_i\times\{y\})=\frac{1}{2}(\sup\sigma_i-\inf\sigma_i)y$ increases in y—and therefore $y^*=d$—and that

$$\varepsilon(\{x\}\times E_Y)=\tfrac{1}{2}x\Delta y, \tag{B.11}$$

increases in x. The error for an efficient $1\to m$ partition is found from (5.1) to be

$$\check{\varepsilon}=\frac{\left|\varphi(b,y^*)-\varphi(a,y^*)\right|}{2m}=\frac{d\Delta x}{2m}. \tag{B.12}$$

To compute the cell boundaries we note that $\tilde{\varphi}^{-1}(z)=z/d$ and from (5.2) we have

$$\sup\sigma_k=\frac{1}{d}\left[ad+2k\frac{d\Delta x}{2m}\right]=a+\frac{k\Delta x}{m}, \tag{B.13}$$

indicating that each cell has the constant width $\Delta x/m$.

It will be efficient to have a nonempty immediate action interval $\sigma_0=[a,x_0)$ only if the smallest possible immediate action error is less than the error without such an interval, viz. $\varepsilon(\{a\}\times E_Y)=\frac{1}{2}a\Delta y\le\check{\varepsilon}$; i.e. a $0\to m$ partition is required when

$$\frac{d\Delta x}{m}>a\Delta y. \tag{B.14}$$

When (B.14) is satisfied, we find $\hat{x}_0$ by equating the immediate action error $\varepsilon_0(\hat{x}_0)=\frac{1}{2}\hat{x}_0\Delta y$ and the message cell error $\varepsilon(\hat{x}_0)=d(b-\hat{x}_0)/2m$, yielding

$$\hat{x}_0=\frac{bd}{m\Delta y+d}. \tag{B.15}$$

Substituting (B.15) into $\varepsilon_0(\hat{x}_0)$, we find that the error for the algorithm is

$$\hat{\varepsilon}=\frac{1}{2}\frac{bd\Delta y}{m\Delta y+d}. \tag{B.16}$$

A calculation similar to that yielding (B.13) shows that the message cells have the constant width

$$\frac{b\Delta y}{m\Delta y+d}\le\frac{bd}{m\Delta y+d}=x_0, \tag{B.17}$$

which we note is weakly less than the width of the immediate action cell.

Summarizing both cases we find the error of any efficient partition for this problem to be $\hat{\varepsilon}$ (respectively $\check{\varepsilon}$) when (B.14) is (respectively is not) satisfied. We will see later in Theorem 8 that we can generalize this result to apply to any multiplicatively separable optimal action function, i.e. one of the form $\varphi(x,y)=f(x)g(y)$, providing neither $f(E_X)$ nor $g(E_Y)$ contain zero as an interior point.

Comparative statics on the communication budget

It is interesting to vary the number of bits n at our disposal—our communication budget—to see the effect upon the structure of and choice of instigator for efficient monologs. First we note that the size of a nonempty immediate action interval for an efficient algorithm must shrink with an increase in n. To see this assume for definiteness that $\varepsilon(\{x\}\times E_Y)$ increases in x. If to the contrary m and $\hat{x}_0$ both increased, then $\varepsilon_0(\hat{x}_0)$ from (5.3) would increase but $\varepsilon(\hat{x}_0)$ from (5.4) would decrease, destroying the required identity $\varepsilon_0(\hat{x}_0(m))\equiv\varepsilon(\hat{x}_0(m))$.

The error $\check{\varepsilon}$ of a $1\rightarrow m$ partition from (5.1) varies inversely with 2^n. The error $\hat{\varepsilon}$ of a $0\rightarrow m$ partition from (5.4) falls more slowly with increases in the communication budget because the numerator increases with m. (This results from the increased length of the interval over which the transmitted bits must distinguish.) We see this in Figure 6, which shows the X-instigator monolog error—with and without an immediate action interval—as a function of the communication budget for $\varphi(x,y)=xy$ on $[0.1,1.1]\times[0,1]$. (The error for a nonempty immediate action interval is bounded below by $\varepsilon(\{0.1\}\times E_Y)=0.05$.)

We have so far studied algorithms which are efficient within the class of X-instigator monologs. Now we ask the question: given a restriction to monolog, who should instigate? Return to the above example of Figure 6. The reader can easily make the parameter interchanges in (B.12), (B.14), (B.15), and (B.16) required to find the solution for a Y-instigator monolog. With one bit an efficient Y-instigator algorithm would result in an immediate action interval of $[0,11/31]$ with an error of $bd\Delta x/2(2\Delta x+b)=11/62$. The best X-instigator monolog would have an error of 11/60. Therefore Y is the preferred instigator and the Y-instigator algorithm is efficient within the class of monologs. Now consider four bits. The best Y-instigator monolog has an immediate action interval of $[0,11/171]$, yielding an error of $11/342=176/5472$. However, with an empty immediate action interval, an X-instigator monolog results in a smaller error of $1/32=171/5472$. So we see that the identity of the optimal instigator can depend on the size of the communication budget.

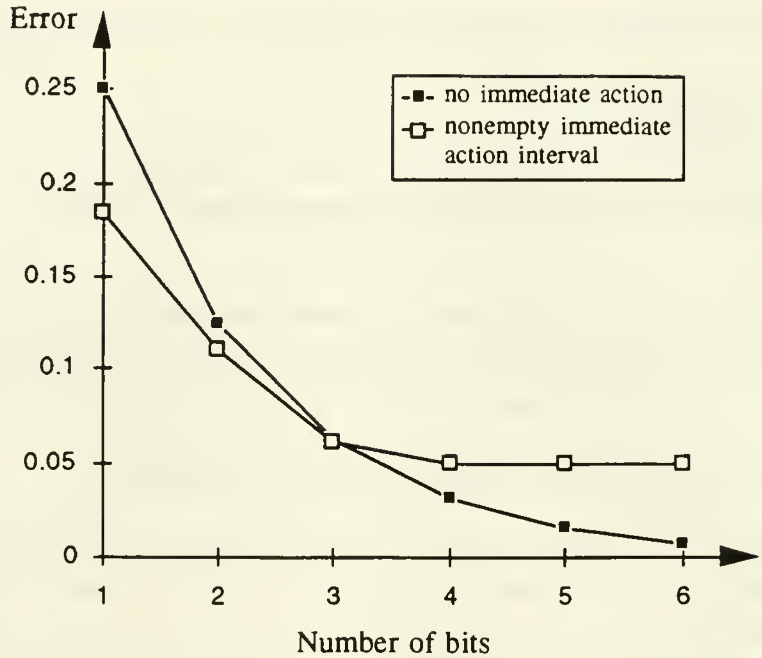

Figure 6: For low n it is efficient to have an immediate action interval.

4. Dialog algorithms

In the previous section we restricted attention to monolog algorithms—those in which any bits sent were transmitted by the instigator. Monologs do not exhaust the algorithmic possibilities, of course. A *dialog* is an algorithm which is not a monolog: for some θ the algorithm requires each partner to send at least one bit. We presented results concerning monologs which are efficient within the class of monologs. We now concern ourselves more generally with the problem of finding algorithms which are efficient within the larger class of all algorithms. We will see that an algorithm which is efficient within the class of monologs need not be efficient within this larger class. (For the remainder of the paper we use "efficient" in this stronger sense.)

Dialog can outperform monolog

Our first example of dialog superiority is extreme—both because it shows a dialog which is infinitely superior to the best monolog and because its optimal action function is somewhat pathological. The second example is more reasonable; its optimal action function is well-behaved and dialog shows only a modest improvement in error relative to monolog.

An extreme example

We display an example in which a dialog outperforms every monolog which has the same communication length. For any communication length greater than unity the dialog error for this optimal action function will be zero and the monolog error will be positive.

Consider the optimal action function

$$\varphi(x,y)=\begin{cases} y, & x=0, \\ 1-x, & y=1, \\ 0, & \text{otherwise}, \end{cases} \tag{C.1}$$

which is graphed in Figure 7. Now consider the following X-instigator dialog algorithm: X sends the message "0" if $x=0$ and sends "1" otherwise. If "0" was sent, Y can immediately choose $\psi=y$ with zero error because he knows X's location precisely. If "1" was sent and $y<1$, then Y can immediately choose $\psi=0$, again with zero error. In the worst case "1" was sent and $y=1$. In this case Y sends any message at all, after which X chooses $\psi=1-x$, again with zero error. So we see that with only two bits a dialog can pin down the value of $\varphi(x,y)$ with zero error even in the worst case.

Now we claim that there is no finite communication length monolog which will determine the value of φ with zero error in the worst-case. Consider any X-instigator monolog. If X immediately acts, either $y<1$ or $y=1$ will yield positive error. If X sends a message, choose as a test case $y=1$. No matter how many bits X sends to Y we can choose an $x>0$ such that Y's estimate of $\varphi(x,y)$ will have positive error, because for $y=1$ the precise determination of $\varphi(x,y)$ requires that the value of x be known precisely. Thanks to the symmetry of the optimal action function we know that this conclusion would hold if Y instigated instead.

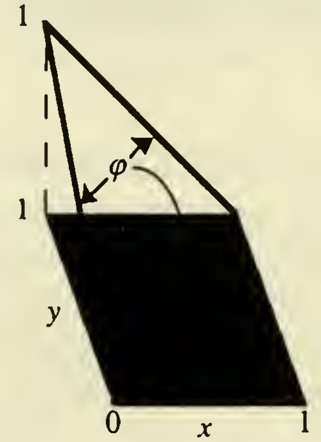

Figure 7: Dialog can be infinitely better than monolog.

A more reasonable example

We return to the optimal action function $\varphi(x,y)=xy$ of Example 3 and choose the domain E to be the unit square $[0,1]\times[0,1]$. We exhibit a two-bit (i.e. $m=4$), X-instigator dialog which outperforms an efficient, two-bit, X-instigator monolog. As the monolog performance benchmark, we calculate, using (B.15$\rightarrow$B.17), that this monolog error is $^1/_{10}$ with an immediate action region over $[0,{}^1/_5)$ and four equally spaced message cells.

In the superior dialog, X immediately acts when $x\in\sigma_0{}^X=[0,0.18468)$. (See Figure 8.) Otherwise X sends the first bit, transmitting "0" when $x\in\sigma_1{}^X=[0.18468,0.63054)$ and transmitting "1" when

$x \in \sigma_2{}^X = [0.63054, 1]$. Y's response depends on X's message. If Y receives "0", he takes control of the next bit. Otherwise, he listens for a second bit to come from X.

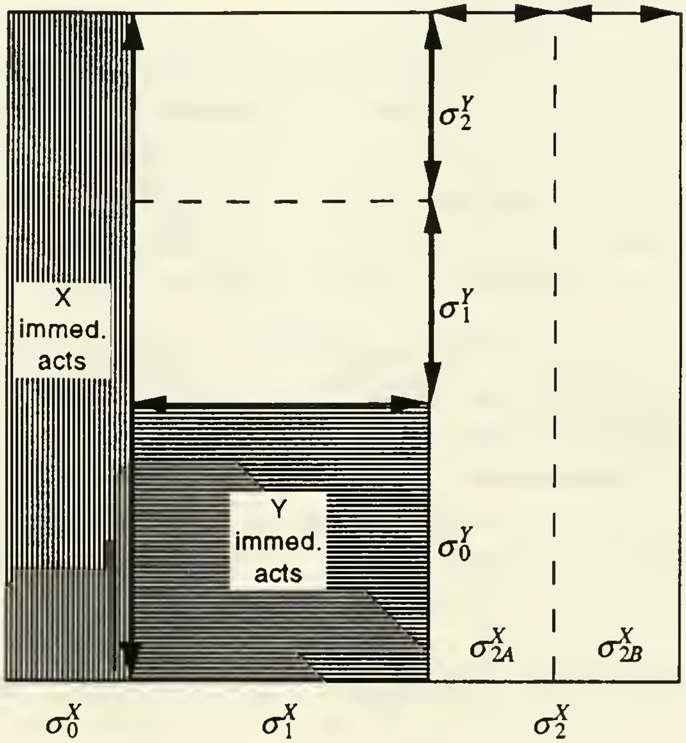

Figure 8: A dialog which outperforms a best monolog.

When Y takes control, he immediately acts if $y \in \sigma_0{}^Y = [0, 0.41421)$. Otherwise he sends "0" as the second bit when $y \in \sigma_1{}^Y = [0.41421, 0.70711)$ and "1" when $y \in \sigma_2{}^Y = [0.70711, 1]$. This then turns the action decision back over to X. Note that when Y receives "0" from X, Y faces a one-bit, Y-instigator, monolog problem on the rectangle $\sigma_1{}^X \times \sigma_2{}^Y$ and solves it the same way we solved the n-bit, X-instigator, monolog problem over an arbitrary rectangle in the previous section.

When $x \in \sigma_2{}^X$ and after X has sent the first bit, she then faces a one-bit, X-instigator monolog problem over the rectangle $\sigma_2{}^X \times E_Y$. The solution to this problem dictates that she use the second bit to further refine Y's knowledge about x into either of the intervals $\sigma_{2A}{}^X = [0.63054, 0.81527)$ or $\sigma_{2B}{}^X = [0.81527, 1]$. Y then takes an action.

The described two-way, two-bit algorithm results in equal-error cells.[1] (The double-headed arrows in Figure 8 indicate the relevant actor's worst-case information set corresponding to each rectangle.) This common error is easily verified to be $0.09234 < {}^1/_{10}$. For example, when X sends "0" and Y responds with "1", X's information set is the line segment $\{x\} \times \sigma_2{}^Y$, where $x \in \sigma_1{}^X$. In the worst-case, $x = \sup \sigma_1{}^X = 0.63054$, and X's best guess, from (A.6), is $\psi = \frac{1}{2}(0.63054)(0.70711 + 1) = 0.53820$, which results, from (A.7), in the error $\frac{1}{2}(0.63054)(1 - 0.70711) = 0.09234$.

---

[1] We are not at this point asserting that the exhibited dialog is efficient; that is not necessary for our claim that it is superior to any efficient monolog.

Monolog is efficient with additive separability

We have seen above that monolog need not be sufficient for efficiency. We now show that whenever the optimal action function is additively separable, monolog *is* efficient—no dialog can do better. By additively separable we mean that the optimal action function is of the form $\varphi(x,y)=f(x)+g(y)$, where $f\colon E_X\to\mathbb{R}$ and $g\colon E_Y\to\mathbb{R}$ are continuous. In order to prove this result we can use the upcoming Theorem 8 to justify restricting attention to optimal action functions of the simpler form

$$\varphi(x,y)=x+y, \tag{C.2}$$

on the closed rectangular region $E=E_X\times E_Y\subset\mathbb{R}^2$.

First we make an observation about this additively separable optimal action function. Let $\sigma_X\subset E_X$ and $\sigma_Y\subset E_Y$. For this φ and for all generalized rectangles $\sigma_X\times\sigma_Y$ the error of an immediate action by X is independent of x and the error of an immediate action by Y is independent of y. To see this we note that

$$\varphi(\{x\}\times\sigma_Y)=\{\varphi(x,y)\colon y\in\sigma_Y\}=\{x+y\colon y\in\sigma_Y\}, \tag{C.3}$$

and therefore

$$\varepsilon(\{x\}\times\sigma_Y)=\tfrac{1}{2}(\sup\{x+y\colon y\in\sigma_Y\}-\inf\{x+y\colon y\in\sigma_Y\})=\tfrac{1}{2}(\sup\sigma_Y-\inf\sigma_Y), \tag{C.4}$$

and similarly for immediate action by Y. As a consequence, every x is a worst-case x and every y is a worst-case y.

Lemma 6A says that it is never efficient to have a nonempty immediate action region. Lemma 6B says that it is never efficient to turn control of any remaining bits over to the other partner, if that partner will use those bits in a monolog. The proof of Theorem 6 then uses an induction argument to establish that it is never efficient for one partner to ever turn control of any remaining bits over to the other partner, whether or not that partner will engage in a monolog from that point forward. Therefore in any efficient algorithm all the bits are sent by the instigator—i.e. only a monolog can be efficient for an additively separable optimal action function. Of course the [illegible] are indep. so [illegible] [illegible] of one agent for the other

Lemma 6A Let $\varphi(x,y)=x+y$ and let $n\geq 1$. Any n-bit, X-instigator algorithm—monolog or dialog—on a generalized rectangle $\sigma_X\times\sigma_Y\subset E$ which has a nonempty immediate action region for the instigator's initial action will be outperformed by some n-bit, Y-instigator monolog whose immediate action region is empty.

Proof Consider an X-instigator, n-bit algorithm where X's control of the first bit generates a partition $\sigma_0\cup\sigma_1\cup\sigma_2=\sigma_X$ where $\sigma_0\neq\varnothing$; i.e. which does have an immediate action region. The error ε^X of this algorithm is at least as large as the error of immediate action $\varepsilon_0{}^X=\tfrac{1}{2}(\sup\sigma_Y-\inf\sigma_Y)$. This algorithm is dominated by an equal-error Y-instigator one-way algorithm which results in an error no larger than[1] $(\sup\sigma_Y-\inf\sigma_Y)/2m<\varepsilon_0{}^X\leq\varepsilon^X$. (In Figure 9a the two-headed arrow corresponds to the oscillation of

---

1 We do not assume that σ_Y is convex; at worst the one-way error is that which would be achieved over the convex hull of σ_Y.

immediate action of the original algorithm. In Figure 9b the shorter two-headed arrows correspond to the oscillation of the optimal action function over X's information set when Y sends a message.) ☺

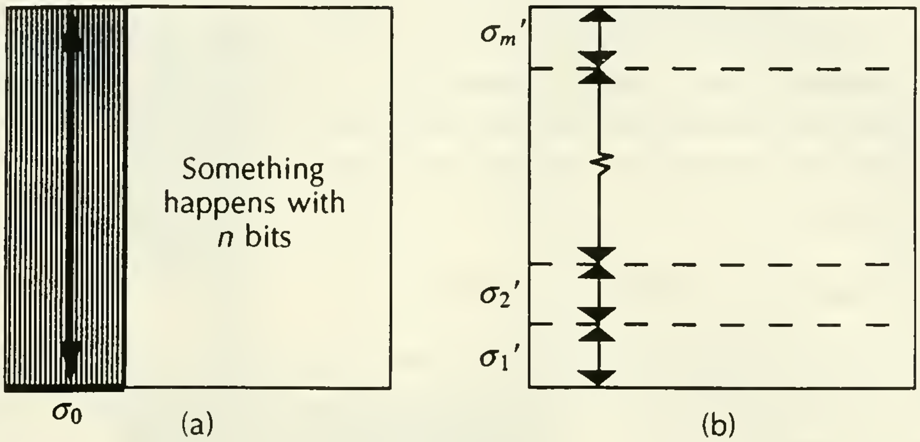

Figure 9: (a) An X-instigator n-bit algorithm with a nonempty immediate action region can be replaced by (b), a lower error Y-instigator, n-bit monolog algorithm which has no immediate action region.

Lemma 6B Let $\varphi(x,y)=x+y$ on E and let $n\geq 2$. Consider any n-bit, X-instigator algorithm on $\sigma_X\times\sigma_Y\subset E$ which for some $x\in\sigma_X$ turns control of the second bit over to Y who then implements an $(n-1)$-bit, monolog algorithm. Then this n-bit, X-instigator algorithm is dominated by an n-bit, Y-instigator, monolog algorithm.

Proof From Lemma 6A we know that without loss of generality we can assume that X's control of the first bit generates a partition $\sigma_1\cup\sigma_2=\sigma_X$. Assume that, in response to some message $i\in\{1,2\}$ from X, Y does take control of the second bit and implements an $(n-1)$-bit monolog on $\sigma_i\times\sigma_Y$ specified by a partition $\{\sigma_{ij}{}^Y\}_{j=1,\ldots,k}$, of σ_Y, where $k=2^{n-1}$. [Lemma 6A assures us that this $(n-1)$-bit, Y-instigator algorithm will not have an immediate action region and therefore $\sigma_{i0}{}^Y=\varnothing$.] For $j\in\{1,\ldots,k\}$, this Y-instigator algorithm results in the cell errors

$$\varepsilon_{ij}=\tfrac{1}{2}(\sup\sigma_{ij}{}^Y-\inf\sigma_{ij}{}^Y), \tag{6B.1}$$

and therefore the error of the original n-bit, X-instigator dialog is bounded below by $\max_j\{\varepsilon_{ij}\}$.

Now define an $(n-1)$-bit, Y-instigator monolog $\hat{\sigma}^Y$ by a partition $\{\hat{\sigma}_j{}^Y\}_{j=1,\ldots,k}$, where $\hat{\sigma}_j{}^Y=\sigma_{ij}{}^Y$ for $j\in\{1,\ldots,k\}$. From (6B.1), $\hat{\varepsilon}_j=\varepsilon_{ij}$ for $j\in\{1,\ldots,k\}$ so that

$$\hat{\varepsilon}^Y=\max_j\{\hat{\varepsilon}_j\}=\max_j\{\varepsilon_{ij}\}\leq\varepsilon^X. \tag{6B.2}$$

This $(n-1)$-bit monolog weakly outperforms the original n-bit dialog; therefore there exists an n-bit monolog which strictly outperforms the dialog. ☺

Theorem 6 The minimum achievable error for n-bits for $\varphi(x,y)=f(x)+g(y)$ on E, where f and g are continuous, can be achieved by a monolog.

Proof We first prove the proposition for the special case of $\varphi(x,y)=x+y$ and then generalize via Theorem 8. Lemma 6B tells us for the case $n=2$ that without loss of generality we can assume that any two-bit algorithm on any generalized rectangle is a monolog. (If a two-bit, X-instigator algorithm were a dialog, then, for some x, Y would send the second bit. However, we know that such an algorithm would not be efficient because it is dominated by some two-bit, Y-instigator monolog.) Now consider $n=3$ and let X be the instigator of an efficient algorithm. From Lemma 6A, X's control of the first bit generates a partition $\sigma_1 \cup \sigma_2 = \sigma_X$. We know from our just previous conclusion for $n=2$ that we can assume that the two-bit algorithm spun-off by each σ_i, $i \in \{1,2\}$, is a monolog. But then Lemma 6B tells us that these two two-bit algorithms must also be X-instigator algorithms; otherwise our algorithm would not be efficient. Therefore we can without loss of generality assume that any efficient three-bit algorithm on any generalized rectangle is a monolog. By induction on n we prove our desired result. From theorem 8 we know that the communication problem for $\varphi(x,y)=f(x)+g(y)$ on E is equivalent to the problem $\hat{\varphi}(x,y)=x+y$ on $\hat{E}=f(E_X)\times g(E_Y)$. Therefore the efficient algorithm for the general additively separable optimal action function is a monolog. ☺

Monolog can be efficient without additive separability

We have seen an example, viz. $\varphi(x,y)=xy$ on $[0,1]\times[0,1]$, in which monolog was insufficient for efficiency. Theorem 6 tells us that monolog is always sufficient for efficiency for additively separable optimal action functions. This raises the question of whether additive separability is not only sufficient but actually necessary for monolog efficiency. We answer this question in the negative by asserting that a best monolog is efficient for $\varphi(x,y)=xy$ when we move the domain to $[1,2]\times[1,2]$.

This demonstration requires more development than required for the earlier example of monolog insufficiency. In that case it was sufficient to exhibit a dialog which outperformed a best monolog; we did not need to claim that the exhibited dialog was efficient. Now, however, in order to show that there exists a monolog which is efficient on this new domain we must prove the nonexistence of a better dialog and therefore we must find a best dialog. Theorem 7 tells us that, for two bits and when the optimal action function is multiplicatively separable, we can greatly restrict our consideration to an extremely small subset of the great variety of conceivable algorithms. This makes the construction of an algorithm to find a best dialog practical. Running this algorithm for $\varphi(x,y)=xy$ on $[1,2]\times[1,2]$ resulted in the finding that every candidate for an efficient dialog was dominated by a monolog.

Before stating the theorem we will develop some notation for discussing two-bit, X-instigator algorithms. X's control of the first bit generates a partition $\sigma_0 \cup \sigma_1 \cup \sigma_2 = E_X = [a,b]$. If the algorithm specifies that when $x \in \sigma_i$, $i \in \{1,2\}$, X maintains control of the second bit as well, then the partition $\sigma_{i0} \cup \sigma_{i1} \cup \sigma_{i2} = \sigma_i$ is generated; i.e. message i leads to a one-bit, X-instigator algorithm on $\sigma_i \times E_Y$. If instead the algorithm specifies that when $x \in \sigma_i$, $i \in \{1,2\}$, Y takes control of the second bit, then the partition $\sigma_{i0}{}^Y \cup \sigma_{i1}{}^Y \cup \sigma_{i2}{}^Y = E_Y = [c, d]$ is generated; i.e. message i leads to a one-bit, Y-instigator algorithm on $\sigma_i \times E_Y$.

Theorem 7 The minimum error achievable by an X-instigator algorithm for the optimal action function $\varphi(x, y) = xy$ on $E \subset \mathbb{R}_+{}^2$ can be achieved by an algorithm of the following form:

① The cells σ_0, σ_1, and σ_2 are convex; $\sigma_0 \ni a$; and $\inf \sigma_1 < \inf \sigma_2$.

② If σ_i, $i \in \{1, 2\}$, leads to a one-bit, X-instigator algorithm, then its immediate action region is empty, i.e. $\sigma_{i0} = \varnothing$.

③ If the one-bit instigator for σ_1 is different than for σ_2, then Y instigates for σ_1 and X instigates for σ_2.

Proof The convexity of σ_0 and the emptiness of σ_{i0} both follow from the monotonicity of $\varepsilon(\{x\} \times E_Y)$. (If $x \in \sigma_{i0}$, then we could without loss of generality incorporate the interval $[a, x]$ into σ_0.) Consider $i \in \{1, 2\}$. If σ_i leads to Y instigating with the remaining bit, then the error over σ_i is

$$\varepsilon_i{}^Y = \max\{\varepsilon_{i0}{}^Y, \varepsilon_{i1}{}^Y, \varepsilon_{i2}{}^Y\},$$

where

$$\varepsilon_{i0}{}^Y = \tfrac{1}{2}\Delta\sigma_i \sup \sigma_{i0}{}^Y,$$

$$\varepsilon_{ij}{}^Y = \tfrac{1}{2}\Delta\sigma_{ij}{}^Y \sup \sigma_i,$$

for $j \in \{1, 2\}$, where $\Delta\sigma \equiv \sup \sigma - \inf \sigma$. This error would be unchanged by the convexification of σ_i. This error would be decreased by shifting σ_i to the left, because this would leave $\Delta\sigma_i$ unchanged and would decrease $\sup \sigma_i$.

If σ_i leads to X instigating again, the error over σ_i is $\varepsilon_i{}^X = \max\{\varepsilon_{i1}{}^X, \varepsilon_{i2}{}^X\}$, where $\varepsilon_{ij}{}^X = \tfrac{1}{2} d \Delta\sigma_{ij}$, $j \in \{1, 2\}$. Each $\varepsilon_{ij}{}^X$ would be unchanged if we convexify and/or translate, either to the right or left, the cell σ_{ij}. If necessary we can convexify σ_i by translating σ_{i1} and σ_{i2} to the right in order to make them adjacent and translating σ_{3-i} to the left, which weakly decreases the error of this cell. This process also establishes ③. ☺

---

5. Separability and Monotonicity

We have thus far imposed monotonicity requirements upon the optimal action function φ and the errors $\varepsilon(\{x\} \times \sigma_Y)$ and $\varepsilon(\sigma_X \times \{y\})$ on the domain E. Theorem 8 allows us to relax these assumptions in many cases. Denote the n-bit efficient communication problem for the optimal action function φ over E by (φ, E). If φ does not satisfy the monotonicity assumptions on E, then we seek to transform (φ, E) into an equivalent problem $(\hat{\varphi}, \hat{E})$ where $\hat{\varphi}$ satisfies the assumptions on $\hat{E}$. The transformed problem would then be amenable to analysis using the methods we have developed thus far.

Consider an optimal action function φ, represented as separable in $f(x)$ and $g(y)$, i.e. such that for all $(x, y) \in E$,

$$\varphi(x, y) = \hat{\varphi}(f(x), g(y)), \tag{D.1}$$

for some function $\hat{\varphi}$. Define $\hat{E}_X = f(E_X)$, $\hat{E}_Y = g(E_Y)$, and $\hat{E} = \hat{E}_X \times \hat{E}_Y$. Then $f\colon E_X \to \mathbb{R}$, $g\colon E_Y \to \mathbb{R}$, and $\hat{\varphi}\colon \hat{E} \to \mathbb{R}$. The separation in (D.1) can be performed trivially for any φ (e.g. let $\hat{\varphi} = \varphi$, $f(x) \equiv x$, and $g(y) \equiv y$); the challenge is to find a nontrivial decomposition which allows the desired transformation of the original communication problem. We require that f and g be continuous so that $\hat{E}_X$ and $\hat{E}_Y$ are intervals and thus $\hat{E}$ is a rectangle.

When we have decomposed the optimal action function à la (D.1) we can conceive of the communication problem facing X and Y in either of two ways: ① to communicate their private information $x \in E_X$ and $y \in E_Y$ in the manner we have studied thus far and then to approximate $\varphi(x, y)$ or ② to communicate $u = f(x) \in \hat{E}_X$ and $v = g(y) \in \hat{E}_Y$ and then to approximate $\hat{\varphi}(u, v)$. Because the optimal action function is sensitive only to the images of x and y under f and g, respectively, these two communication problems are equivalent. Theorem 8 justifies this intuitive argument that the error of the communication problem (φ, E) is equal to the error of the problem $(\hat{\varphi}, \hat{E})$.

Theorem 8 Consider a separable representation of φ as in (D.1). Then the error of any efficient algorithm R for the communication problem (φ, E) is equal to the error of any efficient algorithm $\hat{R}$ for the communication problem $(\hat{\varphi}, \hat{E})$.

Proof An algorithm defines for each partner an operation (which may be null, an action, or the transmission of a message) to be performed at each time t as a function of that partner's coordinate and the history h_t of communication up until that time. Specifically, it defines a sequence of functions $\{\alpha_t(h_t, x)\}_{t=0,\ldots,m}$ for X and $\{\beta_t(h_t, y)\}_{t=0,\ldots,m}$ for Y such that at any time t, and given the history h_t, X performs the operation $a_t = \alpha_t(h_t, x)$ and Y performs the operation $b_t = \beta_t(h_t, y)$.

We construct the sequence $\{\hat{\alpha}_t(h_t, x)\}_{t=0,\ldots,m}$ of operation functions for X for the algorithm $\hat{R}$ from the original sequence for R in the following way. For each $u \in \hat{E}_X$, arbitrarily choose some element of its inverse image,[1]

$$\mu(u) = \operatorname{typ} f^{-1}(u). \tag{8.2}$$

For each t, each $u \in \hat{E}_X$, and each h_t, set

$$\hat{\alpha}_t(h_t, u) = \alpha_t(h_t, \mu(u)). \tag{8.3}$$

---

[1] The function name typ might be interpreted as "typical" or "t[ake] y[our] p[ick]." We use it because other techniques, e.g. taking the infimum or the minimum, can fail. (The infimum may not belong to the set; the minimum may not exist.) It is, of course, crucial that typ truly be a function, i.e. uniquely defined. The only other important property of $\operatorname{typ} S$, for $S \neq \varnothing$, is that $\operatorname{typ} S \in S$.

Similarly, for each $v \in \hat{E}_Y$, let

$$\gamma(v) = \operatorname{typ} g^{-1}(v). \tag{8.4}$$

For each t, each $v \in \hat{E}_Y$, and each h_t, set

$$\hat{\beta}_t(h_t, v) = \beta_t(h_t, \gamma(v)). \tag{8.5}$$

Denote the action resulting from R for $(x, y) \in E$ by $\psi(x, y)$, which results in an error $\varepsilon(x, y)$. Denote the action resulting from $\hat{R}$ for $(u, v) \in \hat{E}$ by $\hat{\psi}(u, v)$, which results in the error $\hat{\varepsilon}(u, v)$. The sets of achieved errors under R and $\hat{R}$, respectively, are

$$\xi = \{\varepsilon(x, y): (x, y) \in E\}, \tag{8.6}$$

$$\hat{\xi} = \{\hat{\varepsilon}(u, v): (u, v) \in \hat{E}\}. \tag{8.7}$$

From (8.3) and (8.5) we see that X and Y will behave at $(u, v) \in \hat{E}$ under algorithm $\hat{R}$ exactly as they would have at $(\mu(u), \gamma(v)) \in E$ under algorithm R. In particular,

$$\hat{\psi}(u, v) = \psi(\mu(u), \gamma(v)), \tag{8.8}$$

and therefore

$$\hat{\varepsilon}(u, v) = \varepsilon(\mu(u), \gamma(v)). \tag{8.9}$$

Because $\mu(u) \in E_X$ and $\gamma(v) \in E_Y$, we see that for all $(u, v) \in \hat{E}$, $\hat{\varepsilon}(u, v) \in \xi$, therefore $\hat{\xi} \subset \xi$, and therefore $\sup \hat{\xi} \leq \sup \xi$. Therefore the error $\hat{R}$ is weakly less than the error of R. However, the error of $\hat{R}$ cannot be strictly less than that of R because R is efficient. (Otherwise the algorithm $\tilde{R}$ on E defined by

$$\tilde{\alpha}(h_t, x) = \hat{\alpha}_t(h_t, f(x)), \tag{8.10}$$

$$\tilde{\beta}(h_t, y) = \hat{\beta}_t(h_t, g(y)), \tag{8.11}$$

would then have an error $\tilde{\varepsilon} = \hat{\varepsilon} < \varepsilon$.) Therefore the error of $\hat{R}$ is equal to the error of R. ☺

In Figure 10 we graph the optimal action function $\varphi(x, y) = \sin 4\pi x + \sin 4\pi y$ on $[0, 1] \times [0, 1]$, which obviously violates our monotonicity assumptions. Letting $f(x) = \sin 4\pi x$, $g(y) = \sin 4\pi y$, and $\hat{\varphi}(u, v) = u + v$, we can then write φ in the separable form (D.1). Therefore we need to solve the problem of $\hat{\varphi}(u, v) = u + v$ on $\hat{E} = f([0, 1]) \times g([0, 1]) = [-1, 1] \times [-1, 1]$. We know from Theorem 6 that monolog is sufficient for efficiency and that we should have an empty immediate action interval. The problem is symmetrical in x and y; we use (5.1) to calculate the efficient error to be $(b - a)/2m = 1/m$ and the cell boundaries to be $u_k = 2k/m - 1$, $k \in M$.

The proof of Theorem 8 [see (8.3)] tells us that we find the partition of E for the original problem from the partition of $\hat{E}$ by finding the inverse image of each cell $\hat{\sigma}_i$, i.e.

$$\sigma_i = f^{-1}(\hat{\sigma}_i) = \frac{1}{4\pi} \sin^{-1}(\hat{\sigma}_i), \tag{D.2}$$

for $i \in M$, where by $\sin^{-1}$ we mean the inverse image and not just the principal value. Figure 11 shows the one-bit case for simplicity. The vertical u axis on the right-hand side shows the convex partition $\hat{\sigma}$ of $\hat{E}$. The two cells of the partition σ for the nonmonotonic optimal action φ each consists of the union of two disjoint intervals, each of which has the same image under f.

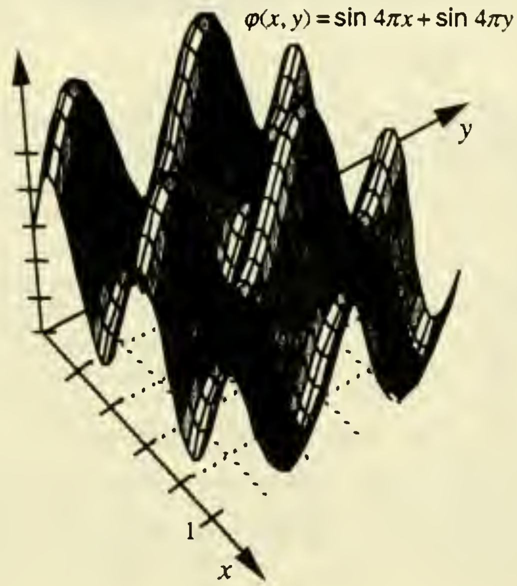

Figure 10: A nonmonotonic but separable function can still be studied.

Similarly, we can study with the techniques we have already developed many multiplicatively separable problems—i.e. where the optimal action function is of the form $\varphi(x,y)=f(x)g(y)$—even when this φ does not itself satisfy the monotonicity assumptions, because $\hat{\varphi}(u,v)=uv$ does satisfy those requirements on any rectangle $\hat{E}$ as long as neither $\hat{E}_X$ nor $\hat{E}_Y$ contain zero as an interior point. (Note that the error of immediate action at u is $\varepsilon(\{u\}\times\hat{\sigma}_Y)=|u|\,\Delta\hat{\sigma}_Y$, which is not monotonic in u over any $\hat{E}_X$ which contains zero in its interior.)

6. Concluding Remarks

We have presented a model of bounded rationality resulting from costly communication in a two-member team context in which the mechanism designer confronted the tradeoff between increased expenditure of resources for communication and increased accuracy of the solution. Our measure of communication emphasized that communication is costly because it is time consuming.

We completely solved the efficient monolog problem for a large class of optimal action functions which satisfied either monotonicity or separability assumptions. This solution included the determination—for any given length of communication—of the optimal instigator, the cell boundaries of the partition, the action decision rules, and the resulting error. Having computed the error as a function of communication length, we have thus characterized the tradeoff between the accuracy and cost of communication. We have shown that there exist optimal action functions such that for some realizations of the private information there will be no communication, even though the right to communicate has already been paid for. The explanation is that remaining silent in optimally chosen, predefined situations sufficiently increases the informativeness of communication in the other cases that this more than compensates for the error of the no-communication actions. We also saw that the identity of the optimal instigator could depend on the size of the communication budget.

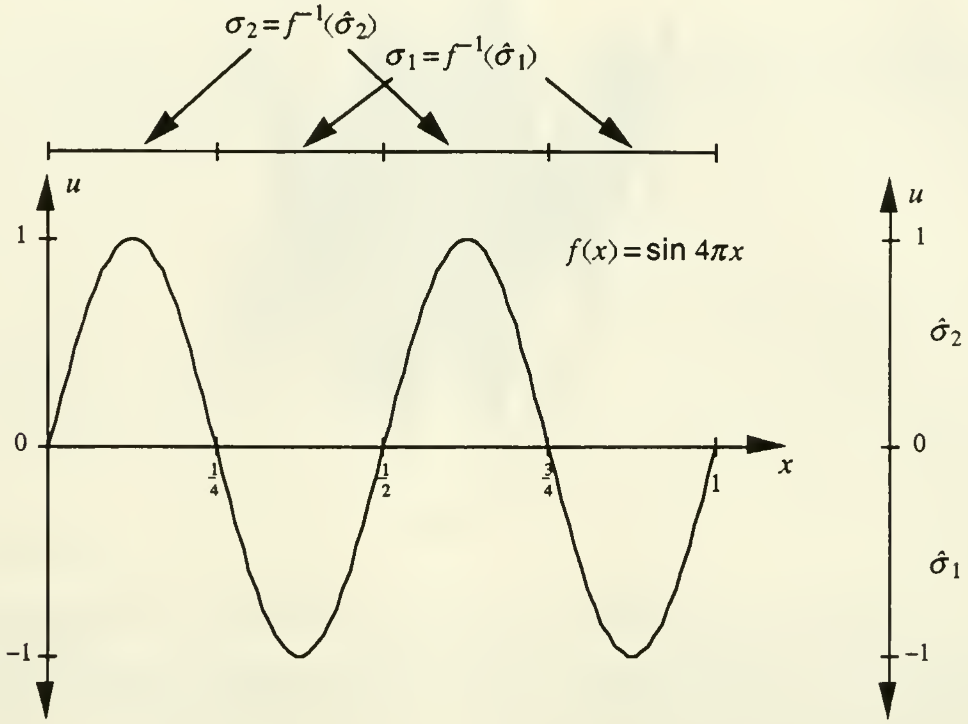

Figure 11: The cell boundaries for the original problem are found from the boundaries of the transformed problem.

We have paid particular attention to the question of whether efficiency requires that both members speak to each other in a dialog rather than that one member engage in a monolog directed to the other. We exhibited examples showing that the answer to this question could in general go either way. If the optimal action function is additively separable, however, we showed that one can restrict attention to monologs without loss of generality. Additive separability corresponds to an independence of the partners' private data.

Admittedly, the information structure in our model—a single real number known by each team member—might seem too austere to realistically pose a significant communication challenge. Our current work-in-progress includes enriching the present framework by endowing each member with private knowledge of a point in a multidimensional space. Our research in this direction has already been informed by both the intuitions and the concrete results of the present work. We have found that interesting multidimensional problems decompose into separate instances of one-dimensional problems of the type we have treated here and thus directly require application of techniques we have developed in this paper.

We see the present paper as one step in a research program dedicated to explaining features of organizational structure which are insufficiently understood when the details of the communication process are ignored.

References

Abelson, Harold [1980] "Lower Bounds on Information Transfer in Distributed Computations," *Journal of the Association for Computing Machinery* **27** 2 (April): 384–392.

Arrow, Kenneth J. [1974] "Limited Knowledge and Economic Analysis," *American Economic Review* **64** 1 (March): 1–10.

Arrow, Kenneth J. and Leonid Hurwicz [1977] "An Optimality Criterion for Decision-making Under Ignorance," in *Studies in Resource Allocation Processes*, eds. Kenneth J. Arrow and Leonid Hurwicz, Cambridge University Press, pp. 463–471.

Dow, James [1991] "Search Decisions with Limited Memory," *Review of Economic Studies* **58** 1 (January): 1–14.

Drèze, Jacques H. and D. de la Vallée Poussin [1971] "A Tâtonnement Process for Public Goods," *Review of Economic Studies* **38**: 133–150.

Green, Jerry R. and Jean-Jacques M. Laffont [1987] "Limited Communication and Incentive Compatibility," in *Information, Incentives, and Economic Mechanisms: Essays in Honor of Leonid Hurwicz*, eds. Theodore Groves, Roy Radner and Stanley Reiter, University of Minnesota Press, pp. 308–329.

Groves, Theodore and John O. Ledyard [1977] "Optimal Allocation of Public Goods: A Solution to the Free Rider Problem," *Econometrica* **45**: 783–809.

Hurwicz, Leonid [1960] "Optimality and Informational Efficiency in Resource Allocation Processes," in *Mathematical Methods in the Social Sciences, 1959: Proceedings of the First Stanford Symposium*, eds. Kenneth J. Arrow, Samuel Karlin and Patrick Suppes, Stanford University Press, pp. 27–46 (Chapter 3).

Jordan, James S. [1982] "The Competitive Mechanism is Dimensionally Efficient Uniquely," *Journal of Economic Theory* **28** 1 (October): 1–18.

Karp, Richard and Antony Ng [forthcoming] *Concrete Complexity.*

Kofman, Fred and James D. Ratliff [1991a] "Monolog vs. Dialog in Costly Bilateral Communication," Interdisciplinary Group on Coordination Theory Working Paper Series, University of California at Berkeley.

Kofman, Fred and James D. Ratliff [1991b] "A Comment on 'Search Decisions with Limited Memory': Waiving the Right to Comparison Shop," mimeo.

Malinvaud, Edmond [1971] "A Planning Approach to the Public Goods Problem," *Swedish Journal of Economics* **73**: 96–112.

Marschak, Jacob and Roy Radner [1972] *Economic Theory of Teams*, Yale University Press.

McAfee, R. Preston and John McMillan [1988] "Search Mechanisms," *Journal of Economic Theory* **44**: 99–123.

Melumad, Nahum, Dilip Mookherjee and Stefan Reichelstein [1989] "Hierarchical Decentralization of Incentive Contracts," Graduate School of Business, Stanford University, mimeo.

Mount, Kenneth R. and Stanley Reiter [1974] "The Informational Size of Message Spaces," *Journal of Economic Theory* **8**: 161–192.

Oniki, Hajime [1986] "The Cost of Communication in Economic Organization: II," in *Uncertainty, Information, and Communication*, eds. Walter P. Heller, Ross M. Starr and David A. Starrett, Cambridge University Press, pp. 191–212.

Osana, Hiroaki [1978] "On the Informational Size of Message Spaces for Resource Allocation Processes," *Journal of Economic Theory* **17**: 66–78.

Tirole, Jean [1988] *The Theory of Industrial Organization*, The MIT Press.

Walker, Mark [1977] "On the Informational Size of Message Spaces," *Journal of Economic Theory* **15**: 366–375.

Yao, Andrew Chi-Chih [1979] "Some Complexity Questions Related to Distributive Computing," *Proceedings of the Eleventh Annual Association of Computing Machinery Symposium on Theory of Computing*: 209–213.

4062 057

Date Due

DEC 19 1991		
FE 27 '92		
NOV. 08 1995		
NOV. [illegible] 1998		

Lib-26-67

MIT LIBRARIES
3 9080 00665066 4