Summary of

Strategic Reliability of Communication Networks

by Michael Suk-Young Chwe

Department of Economics, University of Chicago

1126 East 59th Street, Chicago, Illinois 60637 USA

chwe@uchicago.edu

http://www.spc.uchicago.edu/~wwwchwe/

September 1995

All methods of communication are to some degree unreliable, and one way organizations deal with this unreliability is to form communication networks. This paper offers a game theoretic analysis of how an organization's choice of network depends both on the available communications technology and the underlying strategic situation the organization faces. Previous studies of network reliability do not model the strategic situation and focus on "technical" criteria such as the probability that a message is successfully delivered. This paper's "strategic reliability" approach shows that when the underlying strategic situation is considered, technical criteria are not always appropriate. The paper develops a notation for modelling communication devices and looks at the choice of optimal network in three "coordinated attack" examples.

Keywords: network, communication, game theory, reliability, protocol.

Strategic Reliability of Communication Networks

Michael Suk-Young Chwe
Department of Economics, University of Chicago
1126 East 59th Street, Chicago, Illinois 60637 USA
chwe@uchicago.edu
http://www.spc.uchicago.edu/˜wwwchwe/

September 1995

Introduction

Anyone who asks a friend to repeat herself, or receives his paycheck in person instead of by mail, knows that all methods of communication are to some degree unreliable. Organizations deal with this in at least three related ways. First, they organize communication networks which are less vulnerable to local failures; this was one of the historical design features of the Internet, for example. Second, they make individual messages interpretable even when there are errors; examples range from simply saying the same thing many times to sophisticated error correcting codes. Third, they develop rules about how to communicate: whom to inform next, when to reconfirm, and whom to double-check with, for example; they might be explicit instructions large bureaucracies or just simple habits in families or small offices. An organization, then, must choose a network configuration, a manner in which messages are sent and received, and rules on how communication takes place. These things together I call an organization's "communication protocol."

This paper has benefitted from the comments and encouragement of Hyeok Jeong, Sherwin Rosen, and José Scheinkman, and participants in the Economic Theory Workshop at the University of Chicago.

This paper shows how an organization's communication protocol can depend on both its available communications technology and the underlying strategic situation it faces. For example, if the available communication technology is very unreliable, the communication protocol might involve lots of redundancy or reconfirmation. Organizations for which miscoordination is disastrous, such as emergency rescue teams or military units, would likely have different protocols than organizations for which miscoordination is just inconvenient.

In this paper, communication technology is modelled with "devices" which can be combined to make larger devices such as networks, and the underlying strategic situation is modelled as a game with incomplete information. A communication protocol is modelled as a device together with strategies of the corresponding communication game. A protocol's feasibility is understood in terms of standard game theoretic equilibrium.

This paper looks at three simple but illustrative "coordinated attack" examples, focusing on the network chosen in an optimal communication protocol. The main conclusion is that summary measures of a network's "technical" capabilities, such as the probability that a message is successfully delivered, are not always good at determining whether a network is optimal or even satisfactory. The choice of network depends profoundly on the underlying strategic situation; hence the term "strategic reliability."

Related research

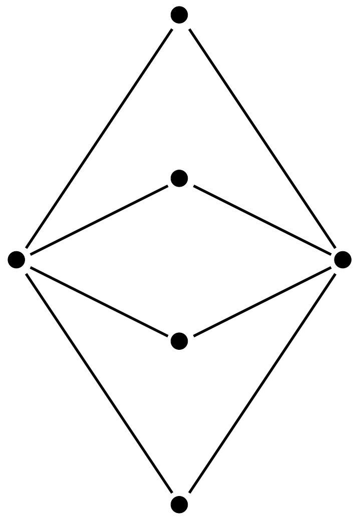
Communication networks have a surprisingly long history; for example, long distance optical telegraph networks with hundreds of relay stations were established well before electric technologies (Wood 1974, Headrick 1991, Holzmann and Pehrson 1994, 1995). The problem of reliability is as old as communication itself; its study recently has become very sophisticated, including the fields of coding theory (Pless 1989) and network reliability theory (Colbourn 1987, Shier 1991).

In modern economics, the amount of communication required by a planned versus a market economy is a classic question which has motivated a rich formal theory (for a survey see Reiter 1986). Communication networks in a firm typically reflect a hierarchical structure; understanding this in terms of incentives (the "principal-agent problem" for instance) has motivated a large literature (Holmstrom and Tirole 1989).

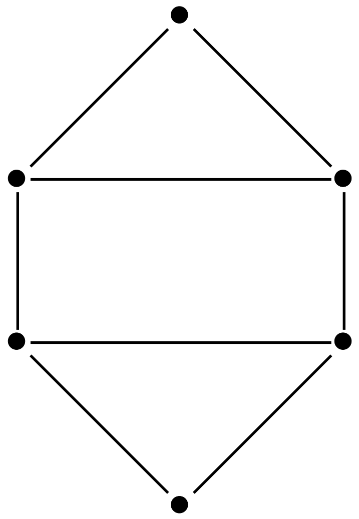
The precursor of much recent research on communication networks is the theory of teams (Marschak and Radner 1972). Recent work includes models based on the assumption that individuals have limited decision-making ability or can specialize in one kind of decision: a firm is a way of decentralizing decision making (Radner and Van Zandt 1992, Radner 1992, Bolton and Dewatripont 1994, Cho and Li 1995, Li 1995). Another line of work sees organizational structure as a way of dealing with individual mistakes in decision making and implicitly assumes that communication is limited (Sah and Stiglitz 1986, Sah and Stiglitz 1988, Sah 1991). Game theoretic models of communication include examples in which noisy communication channels are beneficial (for a survey see Myerson 1991; also Rodes 1995). Finally, there are several models in which there are either explicit costs of sending messages or the set of possible messages are limited (for example Green and Laffont 1987, Townsend 1987, Melumad, Mookherjee, Reichelstein 1992, Prescott 1995).

The research which is most interesting in comparison to this paper is associated with the term "network reliability" (Colbourn 1987, Shier 1991; for an axiomatic game theoretic approach, see McLean and Blair 1991). An organization's communication network is represented mathematically as a graph: each node represents a person, computer, or location, and each edge represents a communication link. A network is connected if for every distinct pair of nodes, there is a sequence of edges which starts at one node and ends at the other. Network reliability can be defined in many ways: one deterministic measure, for example, is the minimum number of edges which when removed disconnect the network.

The most common probabilistic measure is to first assume that each edge has some probability of being operational and then determine the probability that the network is connected. For example, say that there are six people and two competing networks (this example is from Colbourn 1987, p. 48):



Network A



Network B

If each communication link has probability p of being operable (that is, a probability $1-p$ of failing), then Network A is connected with probability $32p^5(1-p)^3+24p^6(1-p)^2+8p^7(1-p)+p^8$ and Network B is connected with probability $30p^5(1-p)^3+25p^6(1-p)^2+8p^7(1-p)+p^8$. Both networks use eight communication links. Interestingly, however, if $p < 2/3$, Network A is more reliable (has a higher probability of being connected). If $p > 2/3$, Network B is more reliable. This is a nice example of how the choice of network topology can depend on the reliability of the individual links.

This paper's "strategic reliability" theory tries to include one thing which network reliability leaves out: the problem the organization faces, or in other words, what the organization actually *does* with its network. The implications of doing this will I hope become clear in the following theory and examples, and will be discussed more fully in the conclusion.

Communication devices

Since an organization's available communication technology and its underlying strategic situation are considered here as two independent influences upon its choice of communication protocol, I model communication technology independently from any game form. The basic element is a "communication device," and is defined to reflect solely technological and not strategic considerations. A device has a sending or "input" end and a receiving or "output" end. People choose one of a possible set of input messages, and then the device outputs one of a possible set of output messages to each person (see also Marschak 1971). Once a communication device is defined, I define how two or more communication devices can be combined to make another communication device.

We assume throughout a finite, nonempty set of individuals $N = \{1, 2, \ldots, n\}$. Throughout this paper, if $A_1, A_2, \ldots, A_n$ are sets, we write $A = \times_{i \in N} A_i$; if $a_1, a_2, \ldots, a_n$ are elements, we write $a = (a_1, \ldots, a_n)$.

Definition. A *communication device* d is a $d = ((X_i)_{i \in N}, (Y_i)_{i \in N}, \sigma)$, where each X_i and Y_i is a nonempty, finite set and the function $\sigma : X \times Y \rightarrow [0, 1]$ satisfies $\sum_{y \in Y} \sigma(x, y) = 1$ for all $x \in X$.

Each person chooses an "input message" $x_i \in X_i$ to send to the device, and then the device sends an "output message" $y_i \in Y_i$ to each person. Given that the people send input messages $x = (x_1, \ldots, x_n) \in X$ to the device, the output message $y = (y_1, \ldots, y_n) \in Y$ is sent back out to the people with probability $\sigma(x, y)$.

The communicative capacity of the device can be represented as the cardinality of X and Y, the number of possible input and output messages. The sets X and Y also represent which people get to send messages and which people receive messages from the device: if X_i is a singleton, this means that person i has effectively no input into the device; similarly, if Y_i

is a singleton, this means that person i does not receive any communication from the device. A message in X_i or Y_i has no literal meaning. The reliability of the device, the probability it will mix up messages, is modelled with σ: for example, if the device is perfectly reliable in the sense that any input message is translated into an particular output message with certainty, then σ will take on values of 0 or 1 only.

Now that we have defined a communication device, we define two ways in which two communication devices can form a new device. The first way to do this is to use the two devices b and c simultaneously.

Definition. Say b and c are two communication devices, where $b = ((X_i^b)_{i \in N}, (Y_i^b)_{i \in N}, \sigma^b)$ and $c = ((X_i^c)_{i \in N}, (Y_i^c)_{i \in N}, \sigma^c)$. Let $d = ((X_i^d)_{i \in N}, (Y_i^d)_{i \in N}, \sigma^d)$, where $X_i^d = X_i^b \times X_i^c$, $Y_i^d = Y_i^b \times Y_i^c$, and $\sigma^d : X^d \to Y^d$ is defined as $\sigma^d(x^d, y^d) = \sigma^d((x^b, x^c), (y^b, y^c)) = \sigma^b(x^b, y^b)\sigma^c(x^c, y^c)$. This d is the communication device in which devices b and c are used *simultaneously*, and we write $d = b \circ c$.

The idea here is that each person simultaneously sends message x_i^b to device b and message x_i^c to device c. Then each person simultaneously receives y_i^b from device b and y_i^c from device c. The two devices b and c operate simultaneously and independently, and thus the probability of output messages (y^b, y^c) given input messages (x^b, x^c) is just the product of the probability of output messages y^b given input messages x^b and the probability of output messages y^c given input messages x^c.

The second way to combine two devices is to use them in sequence.

Definition. Say b and c are two communication devices, where $b = ((X_i^b)_{i \in N}, (Y_i^b)_{i \in N}, \sigma^b)$ and $c = ((X_i^c)_{i \in N}, (Y_i^c)_{i \in N}, \sigma^c)$. Let $d = ((X_i^d)_{i \in N}, (Y_i^d)_{i \in N}, \sigma^d)$, where $X_i^d = X_i^b \times \{f_i : Y_i^b \to X_i^c\}$, $Y_i^d = Y_i^b \times Y_i^c$, and $\sigma^d : X^d \to Y^d$, where $\sigma^d(x^d, y^d) = \sigma^d((x^b, f), (y^b, y^c)) =$

$\sigma^b(x^b, y^b)\sigma^c(f(y^b), y^c)$. This d is the communication device in which devices b and c are used *sequentially* (first b, and then c), and we write $d = b \triangleleft c$.

The idea here is that each person sends input message x_i^b to the first device, device b. Then each person receives output message y_i^b from the first device. Given this output message, person i then decides which input message x_i^c to send to the second device, device c. Finally, each person gets output message y_i^c. These two steps of communication can be thought of as happening in one step: each person chooses input message (x_i^b, f_i), where f_i is a function which indicates which input message x_i^c to send to the second device given the output message received from the first device y_i^b. Then, each person receives the output message (y_i^b, y_i^c). The key to this representation, similar to the representation of an extensive form game as a strategic form game, is that person i's decision of what x_i^c to send given y_i^b can be represented as a contingent rule f_i. Note that although the set $X_i^d = X_i^b \times \{f_i : Y_i^b \to X_i^c\}$ might be very large, it is finite because X_i^b, Y_i^b, and X_c^i are all finite.

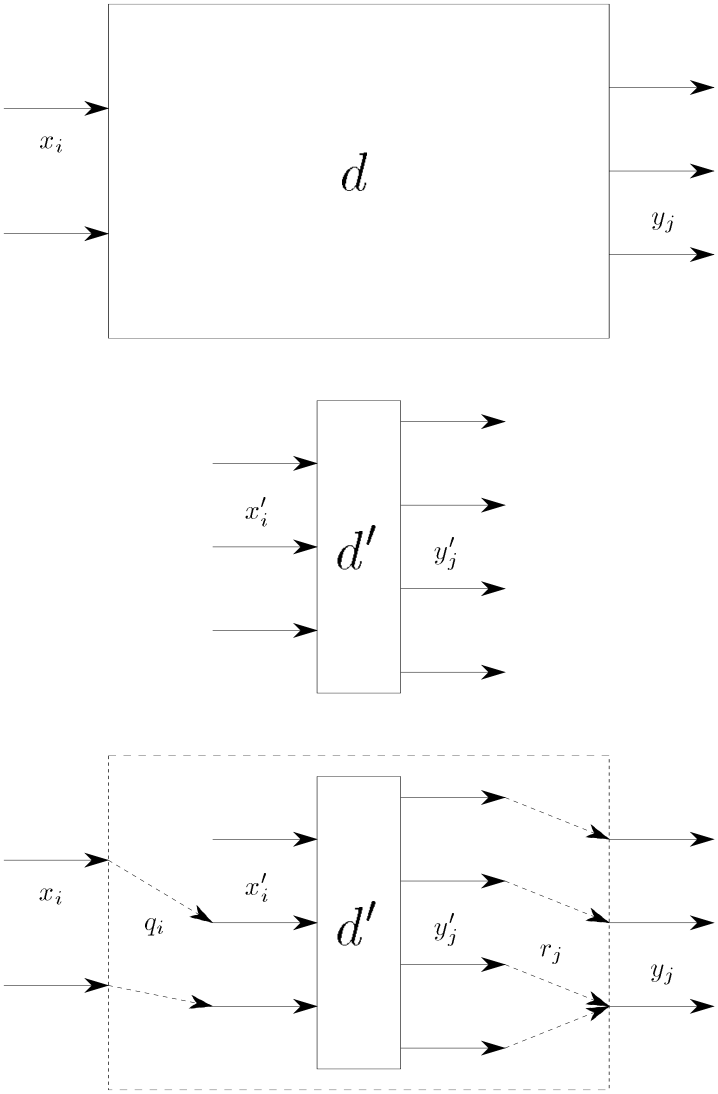
Starting with "primitive" devices, then, more complicated devices can be generated with the binary operations $\circ$ and $\triangleleft$ (this approach is inspired by Shier 1991). We can also define what it means for one device to be "smaller" than another.

Definition. Say $d = ((X_i)_{i \in N}, (Y_i)_{i \in N}, \sigma)$ and $d' = ((X_i')_{i \in N}, (Y_i')_{i \in N}, \sigma')$ are two communication devices. Say there are functions $q_i : X_i \to X_i'$ and $r_i : Y_i' \to Y_i$ such that for all $x \in X$ and $y \in Y$,

$$\sigma(x, y) = \sum_{y' \in r^{-1}(\{y\})} \sigma'(q(x), y')$$

where $r^{-1}(\{y\}) = \{y' \in Y' : r(y) = y'\}$ (when $r^{-1}(\{y\}) = \emptyset$, the sum is 0) and $q : X \to X'$ and $r : Y' \to Y$ are defined as $q(x) = (q_1(x_1), \ldots, q_n(x_n))$ and $r(y') = (r_1(y_1'), \ldots, r_n(y_n'))$. Then we say that d is *smaller than* d' and we write $d \preceq d'$.

This definition is illustrated in the figure below. One can think about the functions q and r as relabelling the inputs and outputs of device d' so that it simulates device d: everything device d can do, device d' can do also. We will make this claim formally later.



The relation $\preceq$ has intuitive properties (the proof of the lemma below is easy and is given as an illustration).

Lemma 1. The relation $\preceq$ is reflexive and transitive; that is, for all devices d, $d \preceq d$, and if b, c, and d are three devices such that $b \preceq c$ and $c \preceq d$, then $b \preceq d$.

Proof. Showing that $d \preceq d$ is obvious. To show $\preceq$ is transitive, let $b \preceq c$ and $c \preceq d$, where $b = ((X_i^b)_{i \in N}, (Y_i^b)_{i \in N}, \sigma^b)$, $c = ((X_i^c)_{i \in N}, (Y_i^c)_{i \in N}, \sigma^c)$, and $d = ((X_i^d)_{i \in N}, (Y_i^d)_{i \in N}, \sigma^d)$. Since $b \preceq c$, by the definition of $\preceq$, $\exists q_i : X_i^b \to X_i^c$ and $\exists r_i : Y_i^c \to Y_i^b$ such that $\sigma^b(x^b, y^b) = \sum_{y^c \in r^{-1}(\{y_b\})} \sigma^c(q(x^b), y^c)$ $\forall x^b \in X^b$, $\forall y^b \in Y^b$. Since $c \preceq d$, by the definition of $\preceq$, $\exists s_i : X_i^c \to X_i^d$ and $\exists t_i : Y_i^d \to Y_i^c$ such that $\sigma^c(x^c, y^c) = \sum_{y^d \in t^{-1}(\{y_c\})} \sigma^d(s(x^c), y^d)$ $\forall x^c \in X^c$, $\forall y^c \in Y^c$. Hence $\sigma^b(x^b, y^b) = \sum_{y^c \in r^{-1}(\{y_b\})} \sum_{y^d \in t^{-1}(\{y_c\})} \sigma^d(s(q(x^b)), y^d)$.

So define $v_i : X_i^b \to X_i^d$ as $v(x_b) = s(q(x_b))$. Define $w : Y_i^d \to Y_i^b$ as $w(y_i^d) = r(t(y_i^d))$. Then $\sum_{y^d \in w^{-1}(\{y_b\})} \sigma^d(v(x^b), y^d) = \sum_{y^d \in w^{-1}(\{y_b\})} \sigma^d(s(q(x^b)), y^d) = \sum_{y^c \in r^{-1}(\{y_b\})} \sum_{y^d \in t^{-1}(\{y_c\})} \sigma^d(s(q(x^b)), y^d)$. But from the above derivation, this is equal to $\sigma^b(x^b, y^b)$ for all $x^b \in X^b$ and $y^b \in Y^b$. Hence $b \preceq d$. □

Now with this definition of "smaller than," we can define what it means for two devices to be equivalent. We then establish some elementary lemmas (presented without proof).

Definition. Say that d and d' are two devices, and say that $d \preceq d'$ and $d' \preceq d$. Then we say that d and d' are *equivalent*, and write $d \approx d'$.

Lemma 2. The relation $\approx$ is an equivalence relation; that is, (1) $b \approx b$, (2) $b \approx c \Rightarrow c \approx b$, and (3) ($b \approx c$ and $c \approx d$) $\Rightarrow b \approx d$ for all devices b, c, d.

Lemma 3. For all devices b, c, d, we have:

(i) $b \circ c \approx c \circ b$;

(ii) $(b \circ c) \circ d \approx b \circ (c \circ d)$;

(iii) $(b \triangleleft c) \triangleleft d \approx b \triangleleft (c \triangleleft d)$.

Lemma 4. Say that b, c, and d are three devices and $b \preceq c$. Then we have:

(i) $b \circ d \preceq c \circ d$;

(ii) $d \circ b \preceq d \circ c$;

(iii) $b \triangleleft d \preceq c \triangleleft d$;

(iv) $d \triangleleft b \preceq d \triangleleft c$.

Lemma 5. For all devices b and c, we have $b \circ c \preceq b \triangleleft c$.

Lemma 6. Say that e is a device which can be written as a finite combination of devices d_i using the binary operations $\circ$ and $\triangleleft$, where the devices d_i appear in order $d_1, d_2, \ldots, d_m$. Then $e \preceq d_1 \triangleleft d_2 \triangleleft \cdots \triangleleft d_m$.

For example, $e = ((d_1 \circ (d_2 \triangleleft d_3)) \circ ((d_4 \circ d_5) \triangleleft d_6 \triangleleft d_7)$. Then $e \preceq d_1 \triangleleft d_2 \triangleleft \cdots \triangleleft d_7$. Linking up individual devices with $\triangleleft$ is always at least as good as linking them up with $\circ$, which makes intuitive sense.

The underlying strategic situation

The organization's underlying strategic situation or underlying game is a game of incomplete information $\Gamma = ((T_i)_{i \in N}, p, (A_i)_{i \in N}, (u_i)_{i \in N})$. Each person $i \in N$ has a nonempty, finite set of possible types T_i, a nonempty, finite set of actions A_i, and a utility function $u_i : T \times A \to \Re$. The prior distribution over types, assumed to be objective for simplicity, is $p : T \to [0, 1]$, where $\sum_{t \in T} p(t) = 1$. To remain compatible with existing definitions (as in Myerson 1991) we assume that $p(t) > 0$ for all $t \in T$: all types occur with nonzero probability.

The communication game and its equilibria

Say that the organization faces an underlying game $\Gamma = ((T_i)_{i\in N}, p, (A_i)_{i\in N}, (u_i)_{i\in N})$ and uses communication device $d = ((X_i)_{i\in N}, (Y_i)_{i\in N}, \sigma)$. The organization uses the communication device in the following way: first, each person finds out her type t_i, and decides which input message x_i to send to the device. The device then sends output messages y_i back to the persons. Then each person, given her type t_i and output message received y_i, decides which action a_i to take. Then the people receive their payoffs.

Given the underlying game Γ and communication device d, the corresponding "communication game" Γ^d is defined as follows. Each person chooses a strategy (g_i, h_i), where $g_i(t_i, x_i)$ is the probability of sending input message x_i given type t_i, and $h_i(t_i, y_i, a_i)$ is the probability of taking action a_i given type t_i and output message y_i. These functions specify how the organization uses the device. Together with the device, they constitute the organization's "communication protocol."

Definition. Say we have an underlying game $\Gamma = ((T_i)_{i\in N}, p, (A_i)_{i\in N}, (u_i)_{i\in N})$ and communication device $d = ((X_i)_{i\in N}, (Y_i)_{i\in N}, \sigma)$. Let $G_i = \{g_i : T_i \times X_i \to [0,1]$ such that $\sum_{x_i \in X_i} g_i(t_i, x_i) = 1 \ \forall t_i \in T_i\}$ and let $H_i = \{h_i : T_i \times Y_i \times A_i \to [0,1]$ such that $\sum_{a_i \in A_i} h_i(t_i, y_i, a_i) = 1 \ \forall t_i \in T_i, \ \forall y_i \in Y_i\}$. Say that $(g_i, h_i) \in G_i \times H_i$ for all $i \in N$. We say (g, h) is a *strategy profile of the communication game* Γ^d. We also call (d, g, h) a *communication protocol* for the game Γ.

Given that each person has strategy (g_i, h_i), what is the "outcome" of the game? Given types t, the probability that input messages x will be sent is $\Pi_{i\in N} g_i(t_i, x_i)$. Given input messages x, the device sends output messages y with probability $\sigma(x, y)$. Given types t and output messages y, the probability that actions a will be taken is $\Pi_{i\in N} h_i(t_i, y_i, a_i)$. Thus the following definitions are immediate.

Definition. Say that (d, g, h) is a communication protocol for the game Γ. Call $\mu[d, g, h](t, a) : T \times A \to [0, 1]$ the *resulting distribution*, where

$$\mu[d, g, h](t, a) = \sum_{x \in X} \sum_{y \in Y} (\Pi_{i \in N} g_i(t_i, x_i))\sigma(x, y)(\Pi_{i \in N} h_i(t_i, y_i, a_i)).$$

One can see from the definition that any correlation between one person's type and another person's action must be accomplished through σ, which represents the communication device.

Now we can make formally the claim that if $d \preceq d'$, everything device d can do, device d' can do also.

Lemma 7. Say that d and d' are two communication devices and that $d \preceq d'$. Say that (d, g, h) is a communication protocol for the game Γ. Then there is a communication protocol (d', g', h') such that $\mu[d, g, h] = \mu[d', g', h']$ (that is, $\mu[d, g, h](t, a) = \mu[d', g', h'](t, a)$ for all $t \in T$ and $a \in A$).

Proof. Notationally complicated but not hard. □

Now that we have defined the communication game's strategies and resulting distribution, we can define payoffs and equilibrium.

Definition. Say that (d, g, h) is a communication protocol for the game Γ. Then person i's *expected utility* $EU_i[d, g, h]$ is defined as

$$EU_i[d, g, h] = \sum_{t \in T} \sum_{a \in A} p(t)\mu(t, a)[d, g, h]u_i(t, a).$$

If $(g, h) \in G \times H$ satisfies the condition that

$$EU_i[d, g, h] \geq EU_i[d, g_i', h_i', (g, h)_{N \setminus \{i\}}] \quad \forall g_i' \in G_i, \forall h_i' \in H_i, \forall i \in N$$

then we say that (d, g, h) is an *incentive compatible* communication protocol for the game Γ. Equivalently, we say that (g, h) is an *equilibrium of the communication game* Γ^d.

The revelation principle

Checking if a given communication protocol (d, g, h) is incentive compatible by using the definition directly can be rather complicated. However, a standard result called the "revelation principle" (see, for example, Myerson 1991, p. 258) allows us to check if (d, g, h) is incentive compatible solely by looking at the resulting distribution $\mu[d, g, h](t, a)$.

We can do this by defining a "mediation plan," a function which for each $t \in T$, yields a probability distribution over the set of actions A. We can then define a device which tries to "implement" this mediation plan: each person i simply sends her own type t_i to a central mediator, and then the central mediator sends to each person i a suggested action a_i. We call this device a "direct revelation" device.

Definition. Say that we have an underlying game $\Gamma = ((T_i)_{i \in N}, p, (A_i)_{i \in N}, (u_i)_{i \in N})$. Let $\eta : T \times A \rightarrow [0, 1]$ be a function which satisfies $\sum_{a \in A} \eta(t, a) = 1$ for all $t \in T$; that is, for each t, $\eta(t, \cdot)$ is a probability distribution over A. We call η a *mediation plan.*

Definition. Given the underlying game Γ and a mediation plan η, define a device $dr^\eta = ((X_i)_{i \in N}, (Y_i)_{i \in N}, \sigma)$, where $X_i = T_i$ and $Y_i = A_i$ for all $i \in N$, and $\sigma(t, a) = \eta(t, a)$. This device dr^η is called a *direct revelation* device.

Definition. Let $g_i^* : T_i \times T_i \rightarrow [0, 1]$ be defined by $g_i^*(t_i, x_i) = 1$ if $x_i = t_i$ and $g_i^*(t_i, x_i) = 0$ otherwise. Let $h_i^* : T_i \times A_i \times A_i \rightarrow [0, 1]$ be defined by $h^*(t_i, y_i, a_i) = 1$ if $a_i = y_i$ and $h^*(t_i, y_i, a_i) = 0$ otherwise. We say that (g^*, h^*) are *truthful and obedient* strategies.

Definition. Given the underlying game Γ and a mediation plan η, we say that η is a *incentive compatible mediation plan* if (dr^η, g^*, h^*) is an incentive compatible protocol for Γ.

In other words, in the direct revelation device dr^η, if each person telling her type truthfully to the mediator and obeying the mediator's suggestion is an equilibrium, then we call

η an incentive compatible mediation plan. The revelation principle goes like this:

Lemma 8. Say we have a game $\Gamma = ((T_i)_{i\in N}, p, (A_i)_{i\in N}, (u_i)_{i\in N})$. If protocol (d, g, h) is incentive compatible, then the resulting distribution $\mu[d, g, h](t, a)$ is an incentive compatible mediation plan.

Proof. See Myerson 1991. □

The point of the revelation principle is that we can check whether a protocol (d, g, h) is incentive compatible solely by checking if the resulting distribution $\mu(t, a)[d, g, h]$ is an incentive compatible mediation plan. Conversely, for every incentive compatible mediation plan, there is a protocol which can "implement" it: namely, the direct revelation device along with the honest and obedient strategies.

If you took all of the possible resulting distributions which result from equilibrium behavior from all possible devices, then, you would simply get the set of incentive compatible mediation plans. In this sense, by specifying a particular device, one does not raise new strategic issues particular to that device only. The question is one of feasibility: given a particular device, which incentive compatible mediation plans can be implemented?

Example: garbling and losing messages

This example is a variation of the "coordinated attack" game in which three people, a sergeant, a private, and a chief, make up a firefighting team which is trying to control a large forest fire. The only incomplete information is whether the fire is weakest at mountain L or at mountain M. The chief is the only one who knows this information, and hence $T_c = \{L, M\}$ (the sergeant's and the private's set of types are singletons $T_s = T_p = \{\Box\}$). Each type is equally likely, and so the objective prior is $p(L) = p(S) = 1/2$. The sergeant and the private each choose among three actions: action l (go to mountain L), action m (go to mountain M), and action n (do nothing). So $A_s = A_p = \{l, m, n\}$. The chief does not take any action, and so his action set is a singleton $A_c = \{\square\}$.

The team is successful only if both the sergeant and the private go to the mountain where the fire is weakest; then everyone gets the success payoff of 1. The sergeant's and the private's payoffs are symmetric: there is a penalty of a for being in the fire alone and there is a penalty of w for going to the wrong location, where the fire is stronger. These penalties are assumed fairly high: $a > 1$ and $w > 1$. Either the sergeant or the private can by doing nothing guarantee herself a payoff of 0. The chief only cares about team success. So the game Γ looks like this, where payoffs are shown as (sergeant, private, chief):

$$\begin{array}{crrr}
 & l & m & n \\
l & 1,1,1 & -a,-a-w,0 & -a,0,0 \\
m & -a-w,-a,0 & -w,-w,0 & -a-w,0,0 \\
n & 0,-a,0 & 0,-a-w,0 & 0,0,0
\end{array}$$

Fire weak on L

$$\begin{array}{crrr}
 & l & m & n \\
l & -w,-w,0 & -a-w,-a,0 & -a-w,0,0 \\
m & -a,-a-w,0 & 1,1,1 & -a,0,0 \\
n & 0,-a-w,0 & 0,-a,0 & 0,0,0
\end{array}$$

Fire weak on M

The team's communication technology consists of two identical devices. Each device is made up of three flags, and hence the possible input messages are the hoisting of one, two, or three flags. With probability ϵ, however, the flags will be obscured by smoke and hence no flags will be seen. There is also a probability δ that the number of flags will be incorrectly counted. In other words, ϵ is the probability that a message will be "lost" and δ is the probability that a message will be "garbled." We define $d_{cs} = ((X_i)_{i \in N}, (Y_i)_{i \in N}, \sigma)$ to be the device in which the chief signals the sergeant: thus $X_c = \{1, 2, 3\}$, $X_s = X_p = \{\square\}$, $Y_s = \{0, 1, 2, 3\}$, and $Y_c = Y_p = \{\square\}$. The function $\sigma : X \to Y$ is defined by

$$\begin{array}{llll}
\sigma(1,0)=\epsilon & \sigma(1,1)=1-2\delta-\epsilon & \sigma(1,2)=\delta & \sigma(1,3)=\delta \\
\sigma(2,0)=\epsilon & \sigma(2,1)=\delta & \sigma(2,2)=1-2\delta-\epsilon & \sigma(2,3)=\delta \\
\sigma(3,0)=\epsilon & \sigma(3,1)=\delta & \sigma(3,2)=\delta & \sigma(3,3)=1-2\delta-\epsilon
\end{array}$$

So the probability that a message "goes through" without a problem is $1-2\delta-\epsilon$. We assume that $1-2\delta-\epsilon>\delta$, that is, the probability that a message goes through is greater than the probability that it gets garbled. The devices d_{cp} (from chief to private) and d_{sp} (from sergeant to private) are defined similarly.

Since the sergeant and private are symmetric, there are basically two competing communication devices: $d_{cs} \circ d_{cp}$, in which the chief informs the sergeant and private separately, and $d_{cs} \triangleleft d_{sp}$, in which the chief informs the sergeant and then the sergeant informs the private. We compare these two mechanisms in terms of what possible values of a, w, ϵ, and δ allow an equilibrium in which there is some probability of success, and also in terms of expected utilities.

First consider the device $d_{cs} \circ d_{cp}$. The first thing to do is to delineate the people's strategies. Recall that each person's strategy is composed of two functions: a probability distribution of what input message to send given her type, and a probability distribution of what action to take given her type, input message, and output message. Things are simpler in this example because of degeneracies: here a strategy profile is specified by the three probability distributions: $g_c(t_c, x_c)$, the chief's probability of sending message x_c given that the type is t_c, $h_s(y_s, a_s)$, the sergeant's probability of taking action a_s given that he received message y_s, and $h_p(y_p, a_p)$, the private's probability of taking action a_p given that he received message y_p.

Consider the strategy profile (g°, h°), defined by:

$$g_c^\circ(L,1)=1 \text{ and } g_c^\circ(L,\cdot)=0 \text{ otherwise;}$$

$$g_c^\circ(M,2)=1 \text{ and } g_c^\circ(M,\cdot)=0 \text{ otherwise;}$$

$h_s^\circ(0,n) = 1$ and $h_s^\circ(0,\cdot) = 0$ otherwise; $h_p^\circ(0,n) = 1$ and $h_p^\circ(0,\cdot) = 0$ otherwise;

$h_s^\circ(1,l) = 1$ and $h_s^\circ(1,\cdot) = 0$ otherwise; $h_p^\circ(1,l) = 1$ and $h_p^\circ(1,\cdot) = 0$ otherwise;

$h_s^\circ(2,m) = 1$ and $h_s^\circ(2,\cdot) = 0$ otherwise; $h_p^\circ(2,m) = 1$ and $h_p^\circ(2,\cdot) = 0$ otherwise;

$h_s^\circ(3,n) = 1$ and $h_s^\circ(3,\cdot) = 0$ otherwise; $h_p^\circ(3,n) = 1$ and $h_p^\circ(3,\cdot) = 0$ otherwise.

In other words, the chief hoists one flag to signal mountain L and two flags to signal mountain M. The sergeant goes to mountain L if he sees one flag, goes to mountain M if he sees two flags, and does nothing if he sees no flags or three flags. The private acts similarly.

For what values of a, w, ϵ, and δ will the protocol $(d_{cs} \circ d_{cp}, g^\circ, h^\circ)$ be incentive compatible? It is easy to see that the chief will signal truthfully. It turns out that our assumption that $a > 1$ and $w > 1$ ensures that the sergeant (and private) will not try to fight the fire after receiving an uninformative message (no flags or three flags). It turns out also that our assumption that successful transmission is more likely than garbled transmission $(1 - 2\delta - \epsilon > \delta)$ ensures that the sergeant (and private) will not go to mountain L after receiving the message to go to mountain M, and vice versa.

So the only "interesting" condition on whether or not $(d_{cs} \circ d_{cp}, g^\circ, h^\circ)$ is incentive compatible involves ensuring that the sergeant (and private) will go to the correct mountain after receiving the signal to do so instead of doing nothing. This condition turns out simply to be that the sergeant and private get a nonnegative expected payoff (since the sergeant and private can always get 0 by doing nothing). Hence $(d_{cs} \circ d_{cp}, g^\circ, h^\circ)$ is incentive compatible if and only if

$$EU_s(\circ) = EU_p(\circ) =$$
$$(1-2\delta-\epsilon)^2 + (-a)(1-2\delta-\epsilon)(2\delta+\epsilon) + (-a-w)\delta(1-\delta) + (-w)\delta^2 \geq 0,$$

where $EU_s(\circ) = EU_s(d_{cs} \circ d_{cp}, g^\circ, h^\circ)$ and $EU_p(\circ) = EU_p(d_{cs} \circ d_{cp}, g^\circ, h^\circ)$ are abbreviations. If this condition is violated, the penalties and risks of miscommunication are together large enough so that the sergeant and private will do nothing even when told to go fight.

One can show without too much difficulty that if (g°, h°) is an equilibrium of the communication game $\Gamma^{d_{cs} \circ d_{cp}}$ (that is, the condition above holds), then there is no other equilibrium which gives any person a higher expected utility: in other words, (g°, h°) is the best equilibrium for everyone.

Now let's look at the competing device $d_{cs} \triangleleft d_{sp}$. Here strategies are specified by four functions: $g_c(t_c, x_c)$, the chief's probability of sending message x_c given that the type is t_c, $g_s(x_s)$, the sergeant's probability of sending input message x_s, $h_s(x_s, y_s, a_s)$, the sergeant's probability of taking action a_s given that he sent message x_s and received message y_s, and $h_p(y_p, a_p)$, the private's probability of taking action a_p given that he received message y_p.

Here, the chief chooses his input message from the set $X_c = \{1, 2, 3\}$, just as before. In this device, however, the sergeant chooses an input message from the set $X_s = \{f_s : \{0, 1, 2, 3\} \rightarrow \{1, 2, 3\}\}$; that is, the sergeant inputs a rule which specifies which message is sent to the private given the message received from the chief. The sergeant and the private each receive an output message from the set $\{0, 1, 2, 3\}$.

Consider the strategy profile $(g^\triangleleft, h^\triangleleft)$, defined by:

$$g_c^\triangleleft(L, 1) = 1 \text{ and } g_c^\triangleleft(L, \cdot) = 0 \text{ otherwise;}$$

$$g_c^\triangleleft(M, 2) = 1 \text{ and } g_c^\triangleleft(M, \cdot) = 0 \text{ otherwise;}$$

$$g_s^\triangleleft(f) = 1 \text{ and } g_s^\triangleleft(\cdot) = 0 \text{ otherwise, where } f \text{ is defined by}$$

$$f(0) = 3,\ f(1) = 1,\ f(2) = 2, \text{ and } f(3) = 3.$$

$$h_s^\triangleleft(x_s, 0, n) = 1 \text{ and } h_s^\triangleleft(x_s, 0, \cdot) = 0 \text{ otherwise, for all } x_s;$$

$$h_s^\triangleleft(x_s, 1, l) = 1 \text{ and } h_s^\triangleleft(x_s, 1, \cdot) = 0 \text{ otherwise, for all } x_s;$$

$$h_s^\triangleleft(x_s, 2, m) = 1 \text{ and } h_s^\triangleleft(x_s, 2, \cdot) = 0 \text{ otherwise, for all } x_s;$$

$$h_s^\triangleleft(x_s, 3, n) = 1 \text{ and } h_s^\triangleleft(x_s, 3, \cdot) = 0 \text{ otherwise, for all } x_s;$$

$$h_p^{\lhd}(0,n) = 1 \text{ and } h_p^{\lhd}(0,\cdot) = 0 \text{ otherwise;}$$
$$h_p^{\lhd}(1,l) = 1 \text{ and } h_p^{\lhd}(1,\cdot) = 0 \text{ otherwise;}$$
$$h_p^{\lhd}(2,m) = 1 \text{ and } h_p^{\lhd}(2,\cdot) = 0 \text{ otherwise;}$$
$$h_p^{\lhd}(3,n) = 1 \text{ and } h_p^{\lhd}(3,\cdot) = 0 \text{ otherwise;}$$

In other words, the chief hoists one flag to signal mountain L and two flags to signal mountain M. If the chief's message is problematic (no flags or three flags), then the sergeant does not attack and notifies the private of the problem by hoisting three flags. If the chief's message is one or two flags, the sergeant relays the message to the private, and goes to mountain L if he sees one flag and mountain M if he sees two. The private goes to mountain L if he sees one flag and goes to mountain M if he sees two, and does nothing if he sees no flags or three flags.

For what values of a, w, ϵ, and δ will the protocol $(d_{cs} \lhd d_{sp}, g^{\lhd}, h^{\lhd})$ be incentive compatible? Like before, the chief will signal truthfully, and our assumptions that $a > 1$ and $w > 1$ and $1 - 2\delta - \epsilon < \delta$ ensure that neither the sergeant or private will try to fight the fire if told not to, and that neither the sergeant or private will go to mountain M if told to go to mountain L, and vice versa. Also, it is easy to see that the sergeant has no reason to deviate from his strategy of how he relays the message to the private.

Like before, the only "interesting" conditions involve making sure the sergeant and private will go fight the fire when told to instead of doing nothing. The conditions for equilibrium turn out again simply to be making sure that both the sergeant and private get a nonnegative expected payoff. Hence $(d_{cs} \lhd d_{sp}, g^{\lhd}, h^{\lhd})$ is incentive compatible if and only if

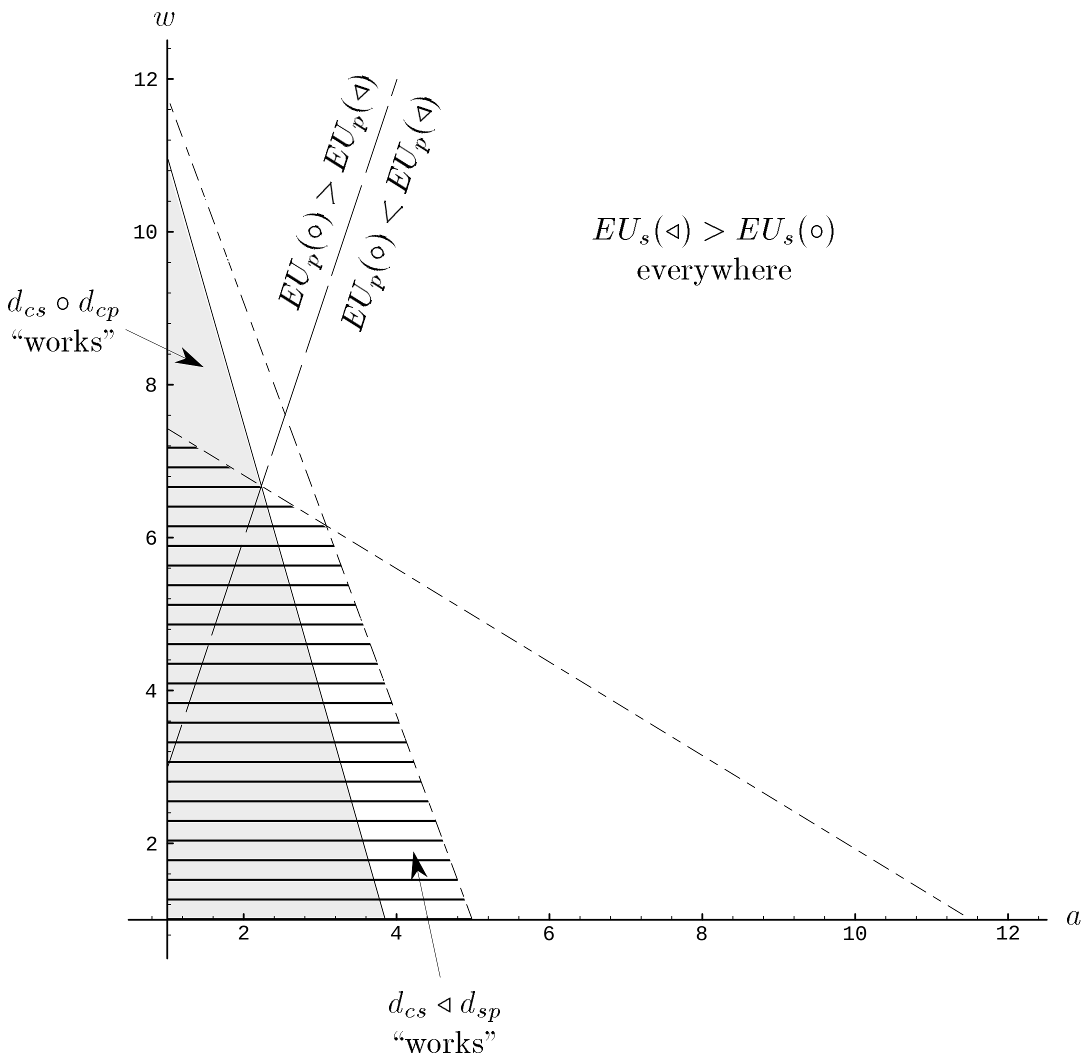
$$EU_s(\lhd) = (1-2\delta-\epsilon)^2 + (-a)(1-2\delta-\epsilon)(2\delta+\epsilon) + (-a-w)\delta(2\delta+\epsilon) + (-w)\delta(1-2\delta-\epsilon) \geq 0$$

and

$$EU_p(\lhd) = (1-2\delta-\epsilon)^2 + (-a)\delta(2\delta+\epsilon) + (-a-w)\delta(1-\delta) + (-w)\delta(1-2\delta-\epsilon) \geq 0,$$

where $EU_s(\triangleleft) = EU_s(d_{cs} \triangleleft d_{sp}, g^\triangleleft, h^\triangleleft)$ and $EU_p(\triangleleft) = EU_p(d_{cs} \triangleleft d_{sp}, g^\triangleleft, h^\triangleleft)$ are abbreviations. Again, one can show without too much difficulty that if $(g^\triangleleft, h^\triangleleft)$ is an equilibrium of the communication game $\Gamma^{d_{cs} \triangleleft d_{sp}}$ (that is, the conditions above hold), then there is no other equilibrium which gives any person a higher expected utility: in other words, $(g^\triangleleft, h^\triangleleft)$ is the best equilibrium for everyone.

Finally, we can compare the two protocols $(d_{cs} \circ d_{cp}, g^\circ, h^\circ)$ and $(d_{cs} \triangleleft d_{sp}, g^\triangleleft, h^\triangleleft)$. For the sake of demonstration, say that $\delta = \epsilon = 0.05$. Then $EU_s(\circ) = EU_p(\circ) = 0.7225 - 0.175a - 0.05w$, $EU_s(\triangleleft) = 0.7225 - 0.135a - 0.05w$, and $EU_p(\triangleleft) = 0.7225 - 0.055a - 0.09w$. In the figure below, the shaded region shows the values of a and w for which $(d_{cs} \circ d_{cp}, g^\circ, h^\circ)$ is incentive compatible ($d_{cs} \circ d_{cp}$ "works"). The striped region shows the values of a and w for which $(d_{cs} \triangleleft d_{sp}, g^\triangleleft, h^\triangleleft)$ is incentive compatible ($d_{cs} \triangleleft d_{sp}$ "works"). Note that in this figure, the origin is at $(1, 1)$ because of our assumption that $a \geq 1$ and $w \geq 1$. The diagonal line splits the quadrant into two regions, one in which the private prefers $d_{cs} \circ d_{cp}$ over $d_{cs} \triangleleft d_{sp}$ and vice versa. Note that the sergeant always prefers $d_{cs} \triangleleft d_{sp}$ over $d_{cs} \circ d_{cp}$ (to abbreviate, we speak of preferences over devices when we really mean preferences over protocols).



From the figure one can see that if a and w are both very large, neither device works: the risks are too great. For a certain set of a and w, both devices work. But for large a and small w, only $d_{cs} \triangleleft d_{sp}$ works: this makes sense intuitively because if the penalty for fighting the fire alone is large, then the sergeant and private are more concerned with getting the same message than getting the right message. Thus the device $d_{cs} \triangleleft d_{sp}$, in which garbling in the single sergeant-to-private link will make someone fight alone, is better than the device $d_{cs} \circ d_{cp}$, in which garbling in either link will make someone fight alone.

For large w and small a, however, only the device $d_{cs} \circ d_{cp}$ works. Here the sergeant and private are more concerned with getting the right message than getting the same message. The key constraint is the private's: in the device $d_{cs} \lhd d_{sp}$, he receives the message "secondhand" and thus his message has two chances of being garbled rather than one.

We can also evaluate the two devices also in terms of the expected utility they give to the team members. The sergeant always gets a higher expected utility from $d_{cs} \lhd d_{sp}$, which makes intuitive sense: he gets the same information as he would in $d_{cs} \circ d_{cp}$, but since the private receives a message from him instead of the chief, it is less likely that the sergeant will fight the fire alone. Whether the private gets a higher expected utility from $d_{cs} \circ d_{cp}$ or $d_{cs} \lhd d_{sp}$ depends on a and w: if the penalty for being in the wrong place w is high, the private prefers getting the information directly in $d_{cs} \circ d_{cp}$; if the penalty for being alone a is high, the private prefers going along with the sergeant in $d_{cs} \lhd d_{sp}$.

The chief, however, is indifferent between the two devices (assuming they both work): both have the same probability of successful attack $(1 - 2\delta - \epsilon)^2$. The devices $d_{cs} \circ d_{cp}$ and $d_{cs} \lhd d_{sp}$ have the same probability of delivering messages correctly; under this "technological" criterion, they are equally good. But people have definite preferences over them, and they have different "specializations": which one the participants prefer, and whether one works and the other does not, depend on the underlying strategic situation.

Example: redundancy and reconfirmation

Two basic ways of dealing with communication unreliability are redundancy and reconfirmation. In redundancy, the same message is sent over multiple independent channels. In reconfirmation, after receiving a message, a confirming message is sent back to the original sender. Here I show how these techniques arise endogenously. This example also illustrates how the binary operations $\circ$ and $\lhd$ and the above lemmas make it possible to consider choosing among all possible technologically feasible communication protocols.

Again, the game Γ considered here involves two people who choose whether to fight a fire, and the weather is either good or bad.

	Fight fire	Do nothing
Fight fire	$1, 1$	$-k_1, 0$
Do nothing	$0, -k_2$	$0, 0$

Good weather

	Fight fire	Do nothing
Fight fire	$-k_1, -k_2$	$-k_1, 0$
Do nothing	$0, -k_2$	$0, 0$

Bad weather

Here either person always can get 0 by doing nothing. If person i fights, and if either he fights alone or fights during bad weather, he will receive a penalty $k_i > 0$. One can think of k_i as the "paranoia level" of person i: the higher it is, the more wary person i is of fighting alone. Only person 1 knows what weather conditions are ($T_1 = \{G, B\}$). The probability of good weather is $p(G) = \gamma$ and probability of bad weather is $p(B) = 1 - \gamma$, where γ is a small number.

The two people have among them m messengers. If a messenger is sent, there is a probability of ϵ that he will not make it to his destination. Of course, if a messenger is not sent, he will certainly not arrive at the destination. Hence define $d_{12}^{\epsilon} = ((X_i)_{i \in N}, (Y_i)_{i \in N}, \sigma)$, where $X_1 = \{1, 0\}$ and $X_2 = \{\Box\}$, $Y_1 = \{\Box\}$ and $Y_2 = \{1, 0\}$, and $\sigma(1,1) = 1 - \epsilon$, $\sigma(1, 0) = \epsilon$, $\sigma(0,0) = 1$, and $\sigma(0, 1) = 0$. The device d_{21}^{ϵ} is defined similarly.

Thus the people can choose any device from the set $D^m = \{d : d$ is formed by combining devices d_{12}^{ϵ} and d_{21}^{ϵ} with the operations $\circ$ and $\triangleleft$, such that there are m or fewer total devices$\}$. For example, we see that $d_{12}^{\epsilon} \circ d_{12}^{\epsilon} \circ d_{12}^{\epsilon} \circ d_{12}^{\epsilon}$ and $d_{12}^{\epsilon} \triangleleft (d_{21}^{\epsilon} \circ d_{21}^{\epsilon}) \triangleleft d_{12}^{\epsilon} \in D^4$. As m increases, the size of D^m increases very quickly; the number of possible devices is in general very large.

For any device $d \in D^m$, there is a trivial equilibrium in which the players never fight regardless of previous communication. So we are interested in incentive compatible protocols (d, g, h), where $d \in D^m$, in which there is a non-zero probability of successful "attack" of the fire. We will look at the Pareto frontier of this set. But first we need some notation and a trivial lemma.

Definition. For a given device d, define $d^{\circ i} = \underbrace{d \circ \cdots \circ d}_{i \text{ times}}$ and $d^{\lhd i} = \underbrace{d \lhd \cdots \lhd d}_{i \text{ times}}$.

Lemma 9. $(d^{\epsilon}_{12})^{\lhd i} \approx (d^{\epsilon}_{12})^{\circ i}$ and $(d^{\epsilon}_{21})^{\lhd i} \approx (d^{\epsilon}_{21})^{\circ i}$.

Proposition 1. Given the game Γ and the set of possible devices D^m , consider the set of incentive compatible protocols (d, g, h) such that $d \in D^m$ and there is a non-zero probability of successful attack. Say that (d, g, h) is on the Pareto frontier of this set. Then either $d = (d^{\epsilon}_{12})^{\circ m}$ or $d = (d^{\epsilon}_{12})^{\circ m_{12}} \lhd (d^{\epsilon}_{21})^{\circ m_{21}}$, where $m_{12} + m_{21} = m$ and $m_{12} \leq m_{21}$.

Sketch of proof. Say (d, g, h) is Pareto efficient among the set of feasible protocols. By Lemmas 6 and 7, without loss of generality, $d = d^{\epsilon}_{12} \lhd d^{\epsilon}_{i_2,3-i_2} \lhd d^{\epsilon}_{i_3,3-i_3} \lhd \cdots \lhd d^{\epsilon}_{i_m,3-i_m}$ (the device surely "starts" with d^{ϵ}_{12} rather than d^{ϵ}_{21} because at the outset person 2 has no information to convey). By Lemma 9, either $d = (d^{\epsilon}_{12})^{\circ j(1)} \lhd (d^{\epsilon}_{21})^{\circ j(2)} \lhd (d^{\epsilon}_{12})^{\circ j(3)} \lhd \cdots \lhd (d^{\epsilon}_{21})^{\circ j(l)}$ or $d = (d^{\epsilon}_{12})^{\circ j(1)} \lhd (d^{\epsilon}_{21})^{\circ j(2)} \lhd (d^{\epsilon}_{12})^{\circ j(3)} \lhd \cdots \lhd (d^{\epsilon}_{12})^{\circ j(l)}$.

In other words, the only devices we need to consider are those in which first person 1 sends $j(1)$ messengers to person 2, person 2 confirms by sending $j(2)$ messengers, person 1 reconfirms by sending $j(3)$ messengers, and so on, ending with either person 1 or person 2 receiving the last group of messengers. It is possible to find the Pareto efficient "sending" strategies g for these devices: person 1 sends $j(1)$ messengers to person 2 if the weather is good and send no messengers if the weather is bad (we call this the first "stage"); if person 2 receives any of these messengers, she sends $j(2)$ messengers back to person 1 to confirm (this is the second stage); if person 1 receives any of these messengers, he sends $j(3)$ messengers to person 2 to reconfirm (the third stage), and so on. Person 1 sends messengers at odd stages and receives them at even stages; person 2 sends messengers at even stages and receives them at odd stages.

The Pareto efficient "action" strategies h are not so immediately obvious. First of all, note that whether one should fight the fire should not depend on the number of messengers which arrive at a given stage. At a given stage, receiving ten messengers does not give a person more information than receiving only one. So whether a person fights or not should only depend on at which stages messengers are received.

Note that because of the "sending" strategies g, if messengers are received at a given stage, then messengers must have been received at every previous stage. Hence, for example, person 1's strategy "Fight if I receive messengers at stages 2 and 8" is equivalent to the strategy "Fight if I receive messengers at all even stages less than or equal to 8." Hence the only strategy profiles we need to consider are: person 1 fights if the weather is good and he receives messengers at all even stages less than or equal to l_1 and person 2 fights if he receives messengers at all odd stages less than or equal to l_2, where $l_1, l_2 \leq l$. If $l_1 = 0$, this means that person 1 attacks based solely on his observation of the weather, without receiving any communication from person 2. For there to be any probability of successful attack, clearly $l_2 \geq 1$. So the set of "strategy profiles" is $\{0, 2, 4, \ldots\} \times \{1, 3, 5, \ldots\}$.

Given the strategy profile (l_1, l_2), the probability of a successful attack (given good weather) is $\prod_{i=1}^{max\{l_1,l_2\}}(1-\epsilon^{j(i)})$. The probability that person i will attack without the other attacking is $\prod_{i=1}^{l_i}(1-\epsilon^{j(i)}) - \prod_{i=1}^{max\{l_1,l_2\}}(1-\epsilon^{j(i)})$ (notationally, we let $\prod_{i=1}^{0} = 1$). Hence person i's expected utility is $\gamma[(1+k_i)\prod_{i=1}^{max\{l_1,l_2\}}(1-\epsilon^{j_i}) - k_i \prod_{i=1}^{l_i}(1-\epsilon^{j_i})]$. Note that if $l_i = max\{l_1, l_2\}$, then person i bears no risk of attacking alone.

Since either person can get a payoff of 0 for sure by never attacking, it is easy to show that a protocol is incentive compatible if and only if it gives a nonnegative expected utility to each person. Say that strategy profile (l_1, l_2) is part of an incentive compatible protocol.

Note that if $l_1 = 0$ and $l_2 \geq 3$, we have $EU_2(l_1, l_2) = (1-\epsilon^{j(l_2)})(1-\epsilon^{j(l_2-1)})EU_2(l_1, l_2-2)$. Also, we can see that $EU_1(l_1, l_2) < EU_1(l_1, l_2 - 2)$. Hence $(l_1, l_2 - 2)$ is incentive compatible and Pareto dominates (l_1, l_2).

Note that if $l_1 \geq 4$ and $l_2 = 1$, we have $EU_1(l_1, l_2) = (1-\epsilon^{j(l_1)})(1-\epsilon^{j(l_1-1)})EU_1(l_1-2, l_2)$. We can also see that $EU_2(l_1, l_2) < EU_2(l_1 - 2, l_2)$. Hence $(l_1 - 2, l_2)$ is incentive compatible and Pareto dominates (l_1, l_2).

If $l_1 \geq 2$ and $l_2 \geq 3$, then $l \geq 3$ and we can write the device d as $d = (d^{\epsilon}_{12})^{\circ j(1)} \triangleleft (d^{\epsilon}_{21})^{\circ j(2)} \triangleleft d'$, where $d' \in D^m$ also. We can write $d' = (d^{\epsilon}_{12})^{\circ j'(1)} \triangleleft (d^{\epsilon}_{21})^{\circ j'(2)} \triangleleft \cdots \triangleleft d^{\epsilon}_{12})^{\circ j'(l-2)}$ or $d' = (d^{\epsilon}_{12})^{\circ j'(1)} \triangleleft (d^{\epsilon}_{21})^{\circ j'(2)} \triangleleft \cdots \triangleleft d^{\epsilon}_{21})^{\circ j'(l-2)}$, where $j'(i) = j(i+2)$. Now if the people use device d' instead of d, and use strategies $(l_1 - 2, l_2 - 2)$ instead of (l_1, l_2), then $EU_i(d', l_1 - 2, l_2 - 2) = [(1 - \epsilon^{j(1)})(1 - \epsilon^{j(2)})]^{-1} EU_i(d, l_1, l_2)$. Hence the protocol $(d', l_1 - 2, l_2 - 2)$ is incentive compatible and Pareto dominates the protocol (d, l_1, l_2).

So the only possible choices for Pareto efficient protocols either have strategies $(l_1, l_2) = (0, 1)$ or $(l_1, l_2) = (2, 1)$. If $(l_1, l_2) = (0, 1)$, then any messengers used past the first confirmation stage are "wasted" (since they don't affect either person's action). Hence the only possible Pareto efficient protocol would use device $(d^{\epsilon}_{12})^{\circ m}$.

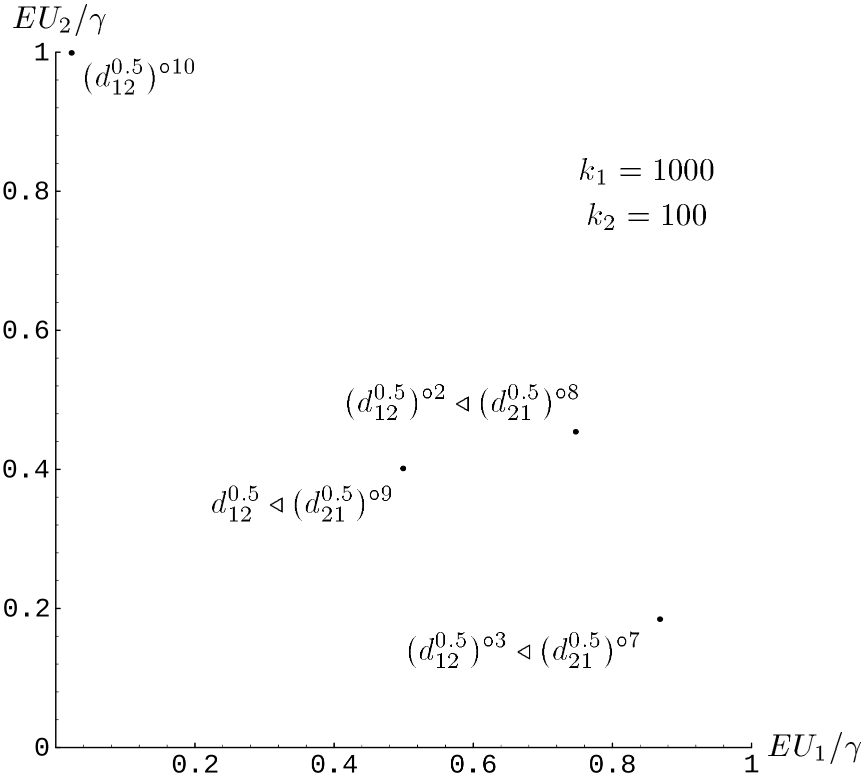
If $(l_1, l_2) = (2, 1)$, then any messengers used past the first two confirmation stages are "wasted." Hence the only possible Pareto efficient protocols use devices $(d^{\epsilon}_{12})^{\circ m_{12}} \triangleleft (d^{\epsilon}_{21})^{\circ m_{21}}$, where $m_{12}+m_{21} = m$. The expected utilities from this protocol are $EU_1 = (1-\epsilon^{m_{12}})(1-\epsilon^{m_{21}})$ and $EU_2 = (1 - \epsilon^{m_{12}})(1 - \epsilon^{m_{21}}) + (1 - \epsilon^{m_{12}})\epsilon^{m_{21}}(-k_2)$. Note that since $EU_2 = (1 - \epsilon^{m_{12}})(1 - \epsilon^{m_{21}}) - k_2\epsilon^{m_{21}} + k_2\epsilon^m$, if $m_{12} > m_{21}$, we can "switch" m_{12} and m_{21} and make person 2 better off while keeping person 1 indifferent. So Pareto efficiency implies that $m_{12} < m_{21}$. □

Hence Pareto efficiency implies either no reconfirmation at all or only one stage of reconfirmation. At first glance, this result seems to directly oppose Rubinstein's (1989) "electronic mail game," in which no amount of reconfirmation makes coordination possible. In his example, two people who are playing a coordinated attack game use computers to communicate (since only person 1 knows where to attack). The communication link between the two computers loses a message with probability ϵ, and the computers are programmed to automatically reconfirm until a message is lost. Each person knows only how many messages his own computer sent. If person 1 sees the number i, for example, he is not sure whether the ith message sent to person 2's computer was lost or the ith message from person 2's computer was lost. The result is that regardless of how small ϵ is or how many reconfirming messages go back and forth, there will never be a coordinated attack.

The reason that Rubinstein's example is nonintuitive is because this device of arbitrarily many (re)^i-confirmations initially seems like a good, perhaps even the best, device available. But as we have shown, (re)^i-confirmations do not help but increasingly hurt. By making the number of potential reconfirmations unbounded, Rubinstein's electronic mail device is actually the worst possible. It should not be surprising that the worst possible device makes coordination impossible. Rubinstein points out that if the computers could be programmed to stop at a fixed number of messages, a coordinated attack is sometimes possible, and it is Pareto efficient to make this fixed number of messages either 1 or 2, which is exactly the result here.

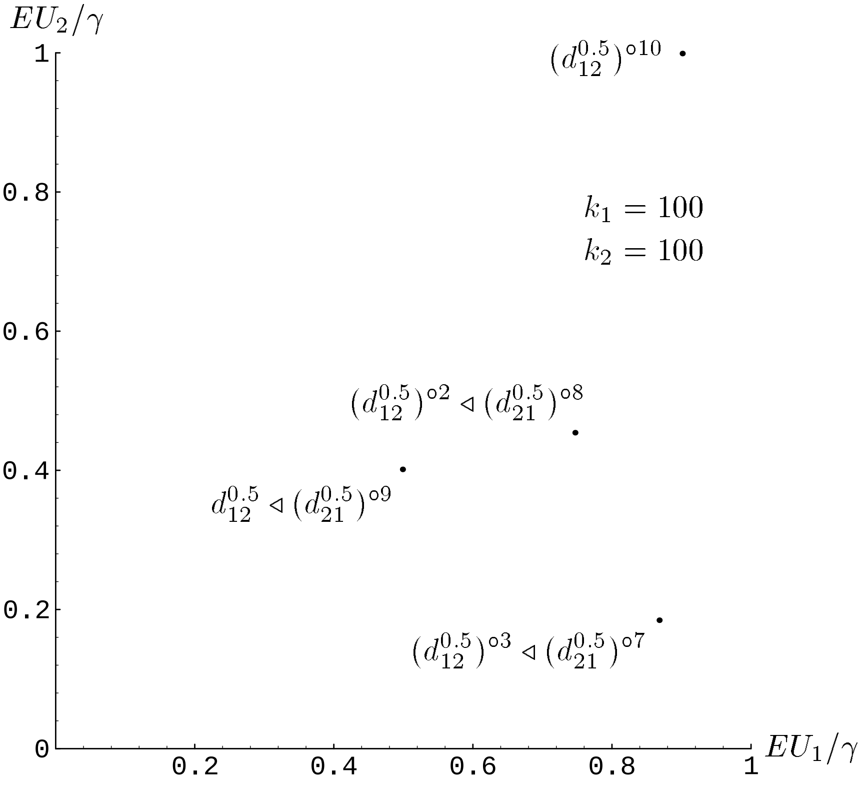
Pareto efficiency is one way to evaluate the feasible protocols. Another way is to look at their probability of successful attack. It is easy to see that $(d_{12}^{\epsilon})^{\circ m}$ gives the highest probability of successful attack, as well as giving the highest expected utility to person 2. If we let h be the largest integer less than or equal to $m/2$, then the second best protocol in terms of probability of successful attack uses device $(d_{12}^{\epsilon})^{\circ h} \triangleleft (d_{21}^{\epsilon})^{\circ m-h}$. The third best is $(d_{12}^{\epsilon})^{\circ h-1} \triangleleft (d_{21}^{\epsilon})^{\circ m-h+1}$, and so on.

As an example, say person 1 has paranoia level $k_1 = 1000$ and person 2 has paranoia level $k_2 = 100$, and say there are $m = 10$ messengers, each with probability $\epsilon = 1/2$ of getting lost. It turns out that there are four incentive compatible protocols, which use devices $(d_{12}^{0.5})^{\circ 10}$, $d_{12}^{0.5} \lhd (d_{21}^{0.5})^{\circ 9}$, $(d_{12}^{0.5})^{\circ 2} \lhd (d_{21}^{0.5})^{\circ 8}$, and $(d_{12}^{0.5})^{\circ 3} \lhd (d_{21}^{0.5})^{\circ 7}$. Their associated expected utilities (divided by γ) are shown in the figure below.



The fact that $(d_{12}^{\epsilon})^{\circ m}$ is always the best for person 2 is clearly shown here. Interestingly, $d_{12}^{0.5} \lhd (d_{21}^{0.5})^{\circ 9}$ is Pareto dominated by $(d_{12}^{0.5})^{\circ 2} \lhd (d_{21}^{0.5})^{\circ 8}$; although $d_{12}^{0.5} \lhd (d_{21}^{0.5})^{\circ 9}$ gives more reassurance to person 2, the increased probability of successful attack of $(d_{12}^{0.5})^{\circ 2} \lhd (d_{21}^{0.5})^{\circ 8}$ (because it is not vulnerable to a failure of a single link) more than compensates. The device $(d_{12}^{0.5})^{\circ 3} \lhd (d_{21}^{0.5})^{\circ 7}$ yields an even higher probability of successful attack, but person 2 does not prefer it.

If we bring person 1's paranoia level down to person 2's ($k_1 = k_2 = 100$), then the incentive compatible protocols are the same four as before, with expected utilities shown below.



Now $(d_{12}^{0.5})^{\circ 10}$ Pareto dominates all of the other feasible protocols. This is naturally the best protocol for person 2. What is different here is that person 1 is now sufficiently unparanoid as to not need any assurance, and thus also favors $(d_{12}^{0.5})^{\circ 10}$ because it gives the highest probability of successful attack. So the set of Pareto efficient protocols depends on k_1 and k_2,and ϵ and m, in interesting ways, and again, the probability the network "works," the probability of successful attack, is only part of the story.

The chain gang example

In her novel *Beloved*, Toni Morrison (1987, p. 110) tells a story about how Paul D, after trying to kill his new owner, was sent to join a chain gang. He and forty-five other men were imprisoned in a row of wooden boxes buried in the earth. Each morning, under armed guard, they chained themselves together, each in turn threading the chain between the loop on his leg irons. Hi Man, the first man on the chain, would shout "Hiiii!" and they would start across the fields to their work.

One day, however, it rained so heavily that work became impossible. The men were locked up in their boxes and in chains all day. As the rain continued for eight straight days, the earth around the boxes turned to mud, and threatened to leak through the boxes and crush them. Suddenly, Paul D felt a firm tug on the chain. "He never figured out how he knew—how anybody did—but he did know—he did—and he took both hands and yanked the length of chain at his left, so the next man would know too." The men tried to escape by pushing through, almost swimming through, the mud.

The chain which bound the forty-six men made escaping a coordination problem. But when the time came, the chain had some crucial advantages. Firstly, the chain was a commitment device ("For one lost, all lost. The chain that held them would save all or none, and Hi Man was the Delivery"). Secondly, the chain was a technology of guiding people and even pulling people through the mud ("Some lost direction and their neighbors, feeling the confused pull of the chain, snatched them around"). Finally, and for us most relevant, the chain was a communication device ("They talked through that chain like Sam Morse, and Great God, they all came up.") The men used this instrument of their domination quite literally as a tool for their liberation.

The tools of this paper show that the men were fortunate, as the chain was in some sense an optimal communication device. Our approach does not match the story exactly:

we model a communication protocol as being decided ahead of time, while this breakout was spontaneous; also, in the story there was a confirming pull in the chain from the other direction which we leave out here for simplicity. This example again shows how it is possible to find the optimal protocols out of the very large set of all possible protocols. (Since I wrote this example, it has become impossible to think of chain gangs only in historical terms; horrifically, Alabama, Arizona, and Florida have reinstituted prison chain gangs (Bragg 1995, Davidson 1995, Dorman 1995, Melone 1995).)

The underlying game Γ is just a n-person version of our previous example: there is uncertainty about whether it is a good time to escape or not, which only person 1 knows ($T_1 = \{G, B\}$ and the prior is $p(G) = \gamma$, $p(B) = 1 - \gamma$, where γ is a small number). Each person chooses his action a_i from the set $A_i = \{e, s\}$ (try to escape or stay put). Person i's utility function is

$$u_i(t, a) = \begin{cases} 1, & \text{if } a = (e, e, \ldots, e) \text{ and } t = G; \\ -k_i, & \text{if } a_i = e \text{ and either } a \neq (e, e, \ldots, e) \text{ or } t = B; \\ 0, & \text{if } a_i = s. \end{cases}$$

If everyone tries to escape and the conditions are good, then everyone gets a payoff of 1. If either conditions are bad or not everyone tries to escape, anyone who tries to escape gets his "paranoia" payoff $-k_i$, where $k_i > 0$. A person who doesn't try to escape always gets utility 0.

Define device d^{ϵ}_{jk} as before: $d^{\epsilon}_{jk} = ((X_i)_{i \in N}, (Y_i)_{i \in N}, \sigma)$, where $X_j = \{1, 0\}$ and $X_i = \{\Box\}$ for $i \neq j$, $Y_k = \{1, 0\}$ and $Y_i = \{\Box\}$ for $i \neq k$, and $\sigma(1, 1) = 1 - \epsilon$, $\sigma(1, 0) = \epsilon$, $\sigma(0, 0) = 1$, and $\sigma(0, 1) = 0$. Here 1 means a pull of the chain, and 0 means no pull of the chain. The idea is that a man might be asleep or might not feel the pull of the chain, and hence there is some probability ϵ that the message does not get through.

Then the chain is simply device $d^{\epsilon}_{12} \triangleleft d^{\epsilon}_{23} \triangleleft \cdots \triangleleft d^{\epsilon}_{n-1,n}$. The strategies they used were g^{*n}, h^{*n}, where $g^{*n}_1 : T_1 \times \{1, 0\} \to [0, 1]$ is defined by $g^{*n}_1(G, 1) = 1$, $g^{*n}_1(G, 0) = 0$,

$g_1^{*n}(B,0) = 0$, and $g_1^{*n}(B,1) = 1$; for $i \in \{2,\ldots,n-1\}$, $g_i^{*n} : \{1,0\} \times \{1,0\} \to [0,1]$ is defined by $g_i^{*n}(1,1) = 1$, $g_i^{*n}(1,0) = 0$, $g_i^{*n}(0,0) = 1$, $g_i^{*n}(0,1) = 0$; g_n^{*n} is degenerate; $h_1 : T_1 \times A_1 \to [0,1]$ is defined by $h_1(G,e) = 1$, $h_1(G,s) = 0$, $h_1(B,e) = 0$, $h_1(B,s) = 1$; and finally for $i = 2,\ldots,n$, $h_i : \{1,0\} \times A_i \to [0,1]$ is defined by $h_i(1,e) = 1$, $h_i(1,s) = 0$, $h_i(0,e) = 0$, $h_i(0,s) = 1$. That is, person 1 pulls the chain if it is a good time to escape, and each person, after feeling the pull of the chain, pulls the chain signalling the next person (since person n doesn't get to send a message to anyone, his g_n^{*n} is degenerate). Person 1 tries to escape if he sees that the conditions are right, and all the other men escape if they feel the chain being pulled.

Altogether $n-1$ "links" are used in the chain. Say that the men use these $n-1$ links to form a different communication device. Could they do any better? With $n-1$ links, they could choose any device in the set $D = \{d : d$ is formed by combining devices d_{ij}^{ϵ} with the operations $\circ$ and $\triangleleft$, such that there are $n-1$ total devices$\}$. The set of possible "chain" devices is $D^* = \{d = d_{i(1),i(2)}^{\epsilon} \triangleleft d_{i(2),i(3)}^{\epsilon} \triangleleft \cdots \triangleleft d_{i(n-1),i(n)}^{\epsilon}$ where $i(1) = 1$ and $\{i(1), i(2), \ldots, i(n)\} = \{1, 2, \ldots, n\}$.

Proposition 2. Given the prison breakout game Γ and the set of possible devices D, consider the set of incentive compatible protocols (d, g, h) such that $d \in D$ and there is a non-zero probability of escape. If (d, g, h) is on the Pareto frontier of this set, then $d \in D^*$, that is, d is a chain device.

Sketch of proof. Let (d, g, h) be incentive compatible, $d \in D$ and there is a non-zero probability of escape. By Lemmas 6 and 7, without loss of generality, $d = d_{i(1),j(1)}^{\epsilon} \triangleleft d_{i(2),j(2)}^{\epsilon} \triangleleft \cdots \triangleleft d_{i(n),j(n)}^{\epsilon}$. Since there are n people and a total of $n-1$ devices, if one person receives more than one message of if person 1 gets a message, then there would be there someone who doesn't get a message, making a non-zero probability of escape impossible. So if we are looking for Pareto efficient protocols, we can safely assume that persons 2 through n

each receive exactly one message. Say that each person gets a message from his predecessor in the network. This predecessor must have a predecessor, and by continuing to follow the "previous" predecessor, we must eventually reach person 1. Thus, we can safely restrict our attention to devices which, if represented as a graph, look like a "tree": the root is at person 1, and every other person has exactly one predecessor. Define person i's "distance from the root" recursively: $r(1) = 0$, and $r(i) = r(j) + 1$, where j is i's predecessor.

Say that person 1 observes that the conditions are good for escaping. Since each other person gets only one message, for there to be a successful escape, all of the devices d^{ϵ}_{ij} have to transmit successfully; hence the probability of successful escape is $(1-\epsilon)^{n-1}$. Person i receives message 1 if all of the $r(i)$ devices between him and person 1 all transmit successfully; hence the probability that person i receives message 1 is $(1-\epsilon)^{r(i)}$. So the probability that person i tries to escape but the escape is not successful is $(1-\epsilon)^{r(i)} - (1-\epsilon)^{n-1}$. Hence person i's (ex ante) expected utility is $EU_i = \gamma[((1-\epsilon)^{r(i)} - (1-\epsilon)^{n-1})(-k_i) + (1-\epsilon)^{n-1}(1)]$, which increases in $r(i)$.

So Pareto efficient devices $d \in D$ are the devices which are Pareto efficient with respect to the root distances $r(i)$. It is easy to show that if $d' \in D \setminus D^*$, there is a $d \in D^*$ which Pareto dominates d' with respect to root distances. □

Note that all devices in D in which each person gets a message yield the same probability of successful escape. A chain device does not make escape any more likely; rather, it is Pareto efficient because it does the best job of assuring each man that when he tries to escape, everyone else will also. In this example, a network's probability of successful operation is not a helpful criterion. Also, a well-known explanation for the existence of hierarchical organization is because hierarchies reduce communication and information processing costs (for example Radner 1992). This example suggests that hierarchies might exist also because of unreliable communication; one might casually observe that the most hierarchical organizations, such as military or emergency rescue teams, are also those which lose the most from miscoordination.

Strategic reliability compared with network reliability

Network reliability theory might be seen as a "macro" approach in contrast with strategic reliability's "micro" approach. First of all, in network reliability, message content, and hence the capacity of communication links and the probability of "garbling," is not considered: a link is either operational and capable of carrying any conceivable message, or not operational and not capable of carrying any message. Also, network reliability assumes that individuals can collectively figure out how to reroute messages in the case of a communication link failure. In strategic reliability, unless it is explicitly specified, a person does not even know whether the message she sends reaches its destination.

Finally, in network reliability, the idea is that the network will be used for a variety of coordinative purposes, like the Internet or the existing telephone network. In network reliability, a network is evaluated on its "technical" aspects, such as the probability that the network will be connected, and thus abstracts away from the needs of the people involved. Strategic reliability precisely specifies the coordination problem as the underlying game, and explores how it influences the choice of the network. By considering people's strategies, strategic reliability analyzes not just the "technical" network but also the "social" protocol.

One thing that network reliability handles nicely is the idea of "node failure": individuals, not just links, can fail. In strategic reliability, it is not immediately clear what it means for a person to "fail": sending no messages, sending random "irrational" messages, and sending sabotaging messages are all possibilities.

The fact that strategic reliability is necessarily more complicated than network reliability might be disheartening, since network reliability quickly becomes very complicated. But this paper shows that it is manageable in some interesting examples.

References

Bolton, Patrick and Mathias Dewatripont. 1994. "The Firm as a Communication Network." *Quarterly Journal of Economics* 109: 809–839.

Bragg, Rick. 1995. "Chain Gangs to Return to Roads of Alabama." *The New York Times,* March 26.

Cho, In-koo and Li Hao. 1995. "Complexity and Network in Repeated Games I: Linear Strategies." Mimeo, University of Chicago.

Colbourn, Charles J. 1987. *The Combinatorics of Network Reliability.* New York: Oxford University Press.

Davidson, Miriam. 1995. "Chain-gang Debate Clangs; Inmates: Work Is 'Humiliating' ". *The Arizona Republic,* June 18.

Dorman, Michael. 1995. "On the Chain Gang; The Debate Rages: Is It Therapy or Slavery?" *New York Newsday,* June 18.

Green, Jerry R. and Jean-Jacques Laffont. 1987. "Limited Communication and Incentive Compatibility." In *Information, Incentives, and Economic Mechanisms,* Theodore Groves, Roy Radner, and Stanley Reiter, editors. Minneapolis, Minnesota: University of Minnesota Press.

Headrick, Daniel R. 1991. *The Invisible Weapon: Telecommunications and International Politics, 1851–1945.* New York: Oxford University Press.

Holzmann, Gerard J. and Björn Pehrson. 1994. "The First Data Networks." *Scientific American,* January 1994.

Holzmann, Gerard J. and Björn Pehrson. 1995. *The Early History of Data Networks.* Los Alamitos, California: IEEE Computer Society Press.

Li, Hao. 1995. *The Organization of Strategies.* Ph.D. dissertation, University of Chicago.

Marschak, Jacob. 1971. "Economics of Information Systems." *Journal of the American Statistical Association* 66: 192–219.

Marschak, Jacob and Roy Radner. 1972. *Economic Theory of Teams.* New Haven, Connecticut: Yale University Press.

McLean, Richard P. and Douglas H. Blair. 1991. "An Axiomatic Characterization of the Reliability Polynomial." In *Reliability of Computer and Communications Networks: Proceedings of a DIMACS Workshop, December 2–4, 1989,* Fred Roberts, Frank Hwang, and Clyde Monma, editors. American Mathematical Society.

Melone, Mary Jo. 1995. "Prison Crew Law Unfettered By Details." *St. Petersburg Times,* June 22.

Melumad, Nahum, Dilip Mookherjee, and Stefan Reichelstein. 1992. "A Theory of Responsibility Centers." *Journal of Accounting and Economics* 15: 445–484.

Morrison, Toni. 1988. *Beloved.* New York: Plume.

Myerson, Roger. 1991. *Game Theory: Analysis of Conflict.* Cambridge, Massachusetts: Harvard University Press.

Pless, Vera. 1989. *Introduction to the Theory of Error-Correcting Codes.* Second edition. New York: Wiley.

Prescott, Edward Simpson. 1995. *A Model of Limited and Costly Communication.* Ph.D. dissertation, University of Chicago.

Radner, Roy. 1992. "Hierarchy: The Economics of Managing." *Journal of Economic Literature* 30: 1382–1415.

Radner, Roy and Timothy Van Zandt. 1992. "Information Processing in Firms and Returns to Scale." *Annales d'Economie et de Statistique* 25–26: 265–298.

Reiter, Stanley. 1986. "Informational Incentive and Performance in the $(\text{New})^2$ Welfare Economics." In *Studies in Mathematical Economics,* Stanley Reiter, editor. Mathematical Association of America.

Rodes, Eduardo Cesar. 1995. *Essays on Communication, Organization, and Game Theory.* Ph.D. dissertation, University of Chicago.

Rubinstein, Ariel. 1989. “The Electronic Mail Game: Strategic Behavior Under ‘Almost Common Knowledge.’ ” *American Economic Review* 79: 385–391.

Sah, Raaj Kumar. 1991. “Fallibility in Human Organizations and Political Systems.” *Journal of Economic Perspectives* 5: 67–88.

Sah, Raaj Kumar and Joseph E. Stiglitz. 1986. “The Architecture of Economic Systems: Hierarchies and Polyarchies.” *American Economic Review* 76: 716–727.

Sah, Raaj Kumar and Joseph E. Stiglitz. 1988. “Committees, Hierarchies and Polyarchies.” *Economic Journal* 98: 451–470.

Shier, Douglas. R. 1991. *Network Reliability and Algebraic Structures.* Oxford: Clarendon Press.

Townsend, Robert M. 1987. “Economic Organization with Limited Communication.” *American Economic Review* 77: 954–971.

Wood, David L. 1974. *A History of Tactical Communication Techniques.* New York: Arno Press.