Worst-Case Interactive Communication I: Two Messages are Almost Optimal*

Alon Orlitsky

AT&T Bell Laboratories

Abstract

X and Y are random variables. Person $P_{\mathcal{X}}$ knows X, Person $P_{\mathcal{Y}}$ knows Y, and both know the joint probability distribution of the pair (X,Y). Using a predetermined protocol, they communicate over a binary, error-free, channel in order for $P_{\mathcal{Y}}$ to learn X. $P_{\mathcal{X}}$ may or may not learn Y. How many information bits must be transmitted (by both persons) in the worst case if only m messages are allowed?

$\hat{C}_1(X|Y)$ is the number of bits required when at most one message is allowed, necessarily from $P_{\mathcal{X}}$ to $P_{\mathcal{Y}}$. $\hat{C}_2(X|Y)$ is the number of bits required when at most two messages are permitted: $P_{\mathcal{Y}}$ transmits a message to $P_{\mathcal{X}}$, then $P_{\mathcal{X}}$ responds with a message to $P_{\mathcal{Y}}$. $\hat{C}_\infty(X|Y)$ is the number of bits required when communication is unrestricted: $P_{\mathcal{X}}$ and $P_{\mathcal{Y}}$ can communicate back and forth.

The maximum reduction in communication achievable via interaction is almost logarithmic. For all (X,Y) pairs, $\hat{C}_\infty(X|Y) \geq \lceil \log \hat{C}_1(X|Y) \rceil + 1$, whereas, for a class of (X,Y) pairs, $\hat{C}_\infty(X|Y) = \lceil \log \hat{C}_1(X|Y) \rceil + 1$. Therefore, $\hat{C}_1(X|Y)$ can be exponentially larger than $\hat{C}_\infty(X|Y)$. Yet $\hat{C}_2(X|Y)$ cannot. With just two messages, the number of bits required is at most four times larger than the minimum: for all (X,Y) pairs, $\hat{C}_2(X|Y) \leq 4\hat{C}_\infty(X|Y) + 3$. This contrasts with communication complexity where, for every m, the number of bits required with m messages can be almost exponentially larger than needed with $m+1$ messages.

Surprisingly, for almost all *sparse* (X,Y) pairs, $P_{\mathcal{Y}}$ who wants to say nothing must transmit almost all the bits in order to achieve the minimum number: $\hat{C}_\infty(X|Y)$. The number of bits transmitted by $P_{\mathcal{Y}}$ can be appreciably reduced only if $P_{\mathcal{X}}$ transmits exponentially more

*Appeared, IEEE Transactions on Information Theory, Sept. 1990, pp. 1111–1127. This version was printed on March 22, 1996.

than $\hat{C}_\infty(X|Y)$ bits.

If the communicators can tolerate ϵ probability of $P_\mathcal{Y}$ not knowing the correct value of X and if randomized protocols are allowed then $4\hat{C}_\infty(X|Y) + 2\log(1/\epsilon)$ bits suffice even without interaction.

1 Introduction and Definitions

Often, when a person conveys information to another, the recipient has some idea about the information sent. Although it is not the recipient's intention to convey any information, he might find that by telling some of his information, the *total* amount of communication can be reduced.

Daily-life examples abound. A sophisticated gossip will ask "Have you heard about ..." before telling a long piece. If the listener has not heard about ..., the amount of data communicated increases by the above phrase and the word "no". If he has, the amount of communication decreases by much more. Although for some cases (corresponding to uninformed individuals) communication increases by a small amount, on the average (assuming a well-gossiped society), communication decreases.

The following is a more concrete example.

Example 1 A league has t teams named $1,\ldots,t$. Every week two random teams play each other. The outcome of the match is announced over the radio. All announcements have the same format: "The match between team I and team J was won by team K" where $1 \le I < J \le t$ and K is either I or J.

$P_{\mathcal{X}}$ and $P_{\mathcal{Y}}$ are avid sports fans. One day, while $P_{\mathcal{Y}}$ listens to a match announcement, $P_{\mathcal{X}}$ grabs the radio from him. Consequently, $P_{\mathcal{Y}}$ hears the first part of the announcement: "The match between team I and team J" and $P_{\mathcal{X}}$ hears the second part: "was won by team K." $P_{\mathcal{X}}$ and $P_{\mathcal{Y}}$ agree that $P_{\mathcal{Y}}$ should know the winner (for it is his radio). They are looking for the most efficient way to do it.

If no interaction is allowed, $P_{\mathcal{X}}$ has to send a single message enabling $P_{\mathcal{Y}}$ to uniquely determine the winner. This message is based solely on the winner (for that is all $P_{\mathcal{X}}$ knows). Let $\sigma(i)$ be the message sent by $P_{\mathcal{X}}$ when he hears "was won by team i." If the messages $\sigma(i)$ and $\sigma(j)$ are the same for some $i \ne j$, then in the event of a match between teams i and j, $P_{\mathcal{Y}}$ cannot tell who the winner is. Also, if $\sigma(i)$ is a prefix of $\sigma(j)$ for $i \ne j$ then in the event of a match between teams i and j $P_{\mathcal{Y}}$ does not know when the message ends. Therefore, the messages $\sigma(1),\ldots,\sigma(t)$ must all be different and none can be a proper prefix of another. Hence the average one-way communication is at least $\log t$ bits and the worst-case one-way communication is at least $\lceil \log t \rceil$ bits. The worst-case lower bound is clearly achievable as the one-way protocol instructing $P_{\mathcal{X}}$ to transmit the binary representation of K (well, the $\lceil \log t \rceil$-bit binary representation of $K-1$ to be precise) requires $\lceil \log t \rceil$ bits in the worst case.

Communication is very different once interaction is allowed. We describe a two-message protocol that requires only $\lceil \log\log t \rceil + 1$ bits in the worst case. $P_{\mathcal{Y}}$ considers the binary representations $I_1, \ldots, I_{\lceil \log t \rceil}$ and $J_1, \ldots, J_{\lceil \log t \rceil}$ of I and J. Since $I \neq J$, there must be a first bit location L where the binary representations differ: $I_L \neq J_L$. Using $\lceil \log\log t \rceil$ bits, $P_{\mathcal{Y}}$ transmits L. $P_{\mathcal{X}}$ responds by transmitting K_L – the L'th bit in the binary representation of the winning team K. The total number of bits exchanged is $\lceil \log\log t \rceil + 1$ bits.

Is there another protocol (using two or more messages) requiring fewer than $\lceil \log\log t \rceil + 1$ bits in the worst case? No. Using the general lower bound proven in Theorem 2, we will see in example 2 that $\lceil \log\log t \rceil + 1$ bits are necessary with two or any number of messages. $\square$

In this paper we explore the scope and limitations of communication reduction achievable via interaction. Our model assumes two *communicators*: an *informant* $P_{\mathcal{X}}$ having a random variable X and a *recipient* $P_{\mathcal{Y}}$ having a, possibly dependent, random variable Y. Both communicators want the recipient, $P_{\mathcal{Y}}$, to learn X with no probability of error, whereas the informant, $P_{\mathcal{X}}$, may or may not learn Y. To that end they alternate in transmitting *messages*: finite sequences of bits. Messages are transmitted over an error-free channel and are determined by an agreed-upon, deterministic, protocol.

A *protocol for* (X, Y) (i.e., a protocol for transmitting X to a person who knows Y) enables communication in a distributed environment where, initially, $P_{\mathcal{X}}$ knows X, $P_{\mathcal{Y}}$ knows Y, and their knowledge about each other's random variables derives from the exchanged messages. It therefore guarantees that the number of messages and bits transmitted is always finite and that the following properties hold.

Separate transmissions: Each message is based on the random variable known to its transmitter and on previous messages.

Implicit termination: When one communicator transmits a message, the other knows when it ends, and when the last message ends, both communicators know that communication has ended.

Correct decision: When communication ends, the recipient, $P_{\mathcal{Y}}$, knows X.

For conciseness of the introduction, we postpone the formal definition of protocols to the next section. Presently, we define the minimum number of bits that must be transmitted in the worst case. The *support set* of a random pair (X,Y) with underlying probability distribution $p(x,y)$ is the set

$$S_{X,Y} \stackrel{\text{def}}{=} \{(x,y) : p(x,y) > 0\}$$

of ordered pairs occurring with positive probability. An *input* is an element of $S_{X,Y}$ viewed as value assignments to X and Y. For every input (x,y) the protocol used, ϕ, determines a finite sequence $\sigma_1(x,y),\ldots,\sigma_{m_\phi(x,y)}(x,y)$ of transmitted messages.[1] Since $P_{\mathcal{X}}$ and $P_{\mathcal{Y}}$ alternate in transmitting these messages, all even numbered messages in this sequence are transmitted by one communicator and all odd numbered messages are transmitted by the other. $m_\phi(x,y)$ is the number of messages exchanged for (x,y). The protocol ϕ is *m-message* if $m_\phi(x,y) \le m$ for all $(x,y) \in S_{X,Y}$. The *length*, $|\sigma|$, of a message σ is the number of bits it contains. The *transmission length* of the input (x,y) is

$$l_\phi(x,y) \stackrel{\text{def}}{=} \sum_{i=1}^{m_\phi(x,y)} |\sigma_i(x,y)| \ ,$$

the number of bits transmitted by both communicators when $X = x$, $Y = y$, and the protocol used is ϕ.

The *worst-case complexity of ϕ* is the supremum, over all inputs, of the number of bits transmitted during the execution of ϕ:

$$\hat{L}_\phi \stackrel{\text{def}}{=} \sup\{l_\phi(x,y) : (x,y) \in S_{X,Y}\} \ .$$

By the definition of protocols, $l_\phi(x,y)$ is always finite. Hence, when $S_{X,Y}$ is finite, so is $\hat{L}_\phi$. If $S_{X,Y}$ is infinite, $\hat{L}_\phi$ may or may not be finite. Many "interesting" random pairs have infinite support sets, yet lend themselves to finite-complexity protocols. One such pair is described in Subsection 4.3, Example 3. The results we prove hold for pairs with finite as well as infinite support sets.

The *m-message complexity of* (X,Y) (i.e., of transmitting X to a person who knows Y) is

$$\hat{C}_m(X|Y) \stackrel{\text{def}}{=} \inf\{\hat{L}_\phi : \phi \text{ is an } m\text{-message protocol for } (X,Y)\} \ ,$$

the number of bits that the two communicators must transmit in the worst case if restricted to protocols that never require more than m messages.

[1] For notational elegance we omit the su(b/per)script ϕ from $\sigma_i(x,y)$. Also, we do not assume that the same number of messages is transmitted for all inputs or that the same communicator transmits the first message for all inputs.

$\hat{C}_1(X|Y)$, the *one-way complexity* of (X,Y), is the number of bits required in the worst case when $P_{\mathcal{Y}}$ cannot transmit to $P_{\mathcal{X}}$, that is, when no *interaction* is allowed. $\hat{C}_2(X|Y)$ is the number of bits required in the worst-case when at most two messages are permitted: $P_{\mathcal{Y}}$ transmits a message reflecting Y, then $P_{\mathcal{X}}$ responds with a message from which $P_{\mathcal{Y}}$ must infer X. Since empty messages are allowed, $\hat{C}_m(X|Y)$ is a decreasing function of m bounded below by 0. We can therefore define $\hat{C}_\infty(X|Y)$, the *unbounded-message complexity of* (X,Y), to be the limit of $\hat{C}_m(X|Y)$ as $m \to \infty$. It is the *minimum* number of bits that must be transmitted for $P_{\mathcal{Y}}$ to know X, even if no restrictions are placed on the number of messages exchanged. A protocol for (X,Y) whose worst-case complexity is $\hat{C}_\infty(X|Y)$ is an *optimal* protocol for (X,Y). In summary, for all random pairs (X,Y):

$$\hat{C}_1(X|Y) \geq \hat{C}_2(X|Y) \geq \hat{C}_3(X|Y) \geq \cdots \geq \hat{C}_\infty(X|Y) .$$

The results obtained in this paper concern the possible discrepancy between these quantities. We now review them, using the league problem of Example 1 for illustration. Casting the example in terms of the general model, $P_{\mathcal{X}}$ knows a winner $X_t \in \{1,\ldots,t\}$ while $P_{\mathcal{Y}}$ knows a match $Y_t \in \{(i,j) : 1 \leq i < j \leq t\}$ and wants to learn X_t. Here, and in the sequel, $X \in \mathcal{X}$ means that the random variable X takes values in the set $\mathcal{X}$. We saw that $\hat{C}_1(X_t|Y_t) = \lceil \log t \rceil$ and that $\hat{C}_2(X_t|Y_t) \leq \lceil \log\log t \rceil + 1$.

Most likely, there is a probability distribution p that determines the match and the winner. However the next section, which formally defines protocols and proves some of their basic properties, also shows that, as we consider only worst-case complexity and do not allow any errors, the precise values of p are irrelevant to $\hat{C}_m(X|Y)$:

Corollary 1 For all (X,Y) pairs and all $m \in \{1,\ldots,\infty\}$, the m-message complexity $\hat{C}_m(X|Y)$ is determined by $S_{X,Y}$, the support set of (X,Y). □

Therefore, the m message complexity of the league problem can be determined although we (implicitly) described only the support set of (X_t,Y_t); since all games are possible and the winner is one of the playing teams,

$$S_{X_t,Y_t} = \{(k,(i,j)) : 1 \leq i < j \leq t \text{ and } k \in \{i,j\}\} .$$

In Section 3 we consider the one-way complexity, $\hat{C}_1(X|Y)$ — the simplest complexity measure to analyze. We define $G(X|Y)$ — the *characteristic hypergraph of* (X,Y). $G(X|Y)$ is simply related to the support set $S_{X,Y}$ and will be instrumental in several subsequent proofs. As a first application we show that it determines the one-way complexity of (X,Y):

Lemma 2 For all (X, Y) pairs,

$$\hat{C}_1(X|Y) = \lceil \log \mathcal{G} \rceil$$

where $\chi(G)$ is the chromatic number of a hypergraph G. □

The proof slightly strengthens a result in Witsenhausen [1] which showed that the minimum number of different messages sent in one-way communication is the chromatic number of a graph equivalent to $G(X|Y)$.

The characteristic hypergraph of the league problem with t teams is K_t, the complete graph on t vertices. This proves, again, that $\hat{C}_1(X_t|Y_t) = \lceil \log t \rceil$ as was shown in Example 1.

We saw that the two-message complexity of the league problem with t teams satisfies: $\hat{C}_2(X_t|Y_t) \le \lceil \log \log t \rceil + 1$. Therefore,

$$\hat{C}_\infty(X_t|Y_t) \le \hat{C}_2(X_t|Y_t) \le \lceil \log \hat{C}_1(X_t|Y_t) \rceil + 1 \ .$$

In Section 4 we prove that this discrepancy is the largest possible. Interaction can never reduce the amount of communication to the logarithm of the one-way complexity:

Theorem 2 For all nontrivial[2] pairs (X, Y):

$$\hat{C}_\infty(X|Y) \ge \lceil \log \hat{C}_1(X|Y) \rceil + 1 \ . \qquad \square$$

Therefore, the two-message protocol described in Example 1 is optimal for the league problem and

$$\hat{C}_2(X_t|Y_t) = \cdots = \hat{C}_\infty(X_t|Y_t) = \lceil \log \log t \rceil + 1 \ .$$

The "almost-logarithmic" reduction achieved for the league problem is typical of a large class of random pairs. Let (X, Y) be a random pair. The *support set* of X is the set

$$S_X \stackrel{\text{def}}{=} \{x : \text{ for some } y,\ (x, y) \in S_{X,Y}\}$$

of possible X values. S_Y, the *support set* of Y, is similarly defined. P_Y's *ambiguity set* when his random variable Y attains the value $y \in S_Y$ is:

$$S_{X|Y}(y) \stackrel{\text{def}}{=} |\{x : (x, y) \in S_{X,Y}\}| \ , \tag{1}$$

the set of possible X values when $Y = y$. His *ambiguity* in that case is

$$\mu_{X|Y}(y) \stackrel{\text{def}}{=} |S_{X|Y}(y)| \ ,$$

[2] A pair (X, Y) is *trivial* if $\hat{C}_1(X|Y) = 0$. That is, X is a deterministic function of Y.

the number of possible X values when $Y = y$. The *maximum ambiguity* of (X, Y) is

$$\hat{\mu}_{X|Y} \stackrel{\text{def}}{=} \sup\{\mu_{X|Y}(y) : y \in S_Y\}, \tag{2}$$

the maximum number of X values possible with any given Y value. In the league problem, the maximum ambiguity is two: for every value known to $P_{\mathcal{Y}}$ (a game), $P_{\mathcal{X}}$ has two possible values (winning teams). Whenever $\hat{\mu}_{X|Y} = 2$, the characteristic hypergraph is a graph, and then, the bound of Theorem 2 is tight:

Theorem 3 For all (X, Y) pairs with $\hat{\mu}_{X|Y} = 2$,

$$\hat{C}_\infty(X|Y) = \lceil \log \hat{C}_1(X|Y) \rceil + 1 \ . \qquad \square$$

Stated differently, for all random pairs with maximum ambiguity two,

$$\hat{C}_1(X|Y) > 2^{\hat{C}_\infty(X|Y)-2}$$

and a similar inequality holds for many other (X, Y) pairs. Therefore, the number of bits required when no interaction is allowed may be exponentially higher than the minimum necessary. Remarkably, Section 5 shows that just two messages (one feedback message) suffice to reduce communication to at most four times the minimum:

Corollary 5 For all (X, Y) pairs,

$$\hat{C}_2(X|Y) \leq 4\hat{C}_\infty(X|Y) + 3 \ . \qquad \square$$

If the maximum ambiguity is two, the corollary can be strengthened still. Just one feedback message suffices to achieve minimum communication:

Corollary 4 For all (X, Y) pairs with $\hat{\mu}_{X|Y} = 2$,

$$\hat{C}_2(X|Y) = \ldots = \hat{C}_\infty(X|Y) = \lceil \log \hat{C}_1(X|Y) \rceil + 1 \ . \qquad \square$$

As a special case we have the league problem. Its maximum ambiguity is two and therefore it has the optimal two-message protocol described in Example 1. We note that for some random pairs two messages are not optimal. A subsequent paper, Orlitsky [2], exhibits a random pair (X, Y) for which $\hat{C}_2(X|Y) \geq 2\hat{C}_\infty(X|Y)$. The largest possible ratio between $\hat{C}_2(X|Y)$ and $\hat{C}_\infty(X|Y)$ is therefore between two and four; its precise value is yet unknown. Later in the introduction we contrast Corollary 5 with the corresponding results in communication complexity.

The two-message protocol of the league problem and many of the protocols providing the upper bound on $\hat{C}_2(X|Y)$ in Section 5 exhibit a peculiar property. $P_{\mathcal{Y}}$, the recipient who wants to say nothing, transmits almost all the bits while the informant, $P_{\mathcal{X}}$, transmits almost none. This does not mean that $P_{\mathcal{Y}}$ must transmit more bits than $P_{\mathcal{X}}$. These random pairs, like all others, have protocols in which $P_{\mathcal{Y}}$ does not transmit at all. Potentially, they could have other optimal protocols in which $P_{\mathcal{Y}}$ transmits only few bits. Surprisingly, Section 6 shows that for almost all (X,Y) pairs with any given maximum ambiguity this is not the case. All optimal or near optimal protocols for these pairs require $P_{\mathcal{Y}}$ to transmit almost $\hat{C}_\infty(X|Y)$ bits, the total number exchanged. The number of bits transmitted by $P_{\mathcal{Y}}$ can be appreciably below $\hat{C}_\infty(X|Y)$ only if $P_{\mathcal{X}}$ transmits exponentially more than $\hat{C}_\infty(X|Y)$ bits. These conclusions are based on the following result.

Theorem 5 Let ϕ be a protocol for a random pair (X,Y). Let $\hat{L}_\phi^{\mathcal{X}}$ denote the maximum, over all inputs, of the number of bits transmitted by $P_{\mathcal{X}}$ under ϕ. Define $\hat{L}_\phi^{\mathcal{Y}}$ similarly. Then

$$\hat{L}_\phi^{\mathcal{X}} \geq \left\lceil \frac{\hat{C}_1(X|Y)}{2^{\hat{L}_\phi^{\mathcal{Y}}}} \right\rceil . \qquad \Box$$

Lastly, in Section 7 we consider randomized protocols. Assuming that unbiased bits can be generated and that the communicators are willing to tolerate ϵ probability of $P_{\mathcal{Y}}$ not knowing the correct value of X, we demonstrate a one-way randomized protocol whose expected transmission for the worst input is only four times larger than $\hat{C}_\infty(X|Y)$.

Corollary 6 For all (X,Y) pairs,

$$\hat{\mathcal{C}}_1^\epsilon(X|Y) \leq 4\hat{C}_\infty(X|Y) + 2\log\frac{1}{\epsilon} ,$$

where $\hat{\mathcal{C}}_1^\epsilon(X|Y)$ is appropriately defined. $\Box$

The results obtained in this paper are related to two existing approaches in the study of communication. One branch of information theory assumes that a communicator wants to convey some random information X to another communicator who knows related information Y. Traditionally, the random pair (X,Y) is a collection of independent identically-distributed (*i.i.d.*) random pairs (X_i,Y_i) and the recipient is willing to tolerate some probability of not knowing X.

Communication complexity literature, on the other hand, assumes that two persons, one knowing X the other knowing Y, communicate in order to determine the value of a binary function $f(X,Y)$. Only one instance ("episode") of the random variable pair is given and the communicators want to know $f(X,Y)$ with no probability of error.

Following El Gamal and Orlitsky [3], we combine the two approaches. As in information theory, we assume that $P_{\mathcal{X}}$ wants to transmit X to $P_{\mathcal{Y}}$. As in communication complexity we assume that only one instance of the random pair is given and that no errors are allowed in determining the value of X.

A few comparisons are in order. The communication complexity problem is more symmetric as f depends on both X and Y. It is therefore natural to expect interaction to reduce communication. Indeed, Papadimitriou and Sipser [4] showed a function for which one-message protocols require $\sqrt{t}/(2\log t)$ transmitted bits while there are two-message protocols for that function requiring only $2\log t+1$ bits. Duris, Galil, and Schnitger [5] showed that for every m, there is a function whose m-message complexity is $\Omega(\sqrt{t}/(m^4\log t))$ but its $(m+1)$-message complexity is only $O(m\log t)$. Hence there can be an almost-exponential gap between m- and $(m+1)$-message complexities. This contrasts with our case. Here $\hat{C}_2(X|Y),\ldots,\hat{C}_\infty(X|Y)$ are all within a factor of four from each other while only $\hat{C}_1(X|Y)$ can be exponentially higher.

To compare with information theory results, consider the related setup of [3]. There, $P_{\mathcal{X}}$ and $P_{\mathcal{Y}}$ want to exchange their information: $P_{\mathcal{Y}}$ wants to learn X and $P_{\mathcal{X}}$ wants to learn Y. It was shown that, *on the average*, at least $H(X|Y)+H(Y|X)$ bits must be transmitted and that $H(X|Y)+H(Y|X)+3\min\{\log\log|S_{\mathcal{X}}|,\log\log|S_{\mathcal{Y}}|\}$ bits suffice for all distributions uniform over their support.[3] This resembles the result of Slepian and Wolf [6] but does not allow any errors in determining X and Y and does not assume that X and Y are collections of i.i.d. random variables. In this paper we are concerned with worst-case complexity, rather than average, and we study the effect of interaction, rather than determine the minimum number of bits that must be transmitted.

2 Protocols and Preliminary Results

Protocols were informally described in the introduction. For conciseness however we left out the details. In Sections 3 and 4 we prove lower bounds on the complexities of probability distributions. These bounds rely on the precise definitions of protocols given in this

[3] $H(X)\stackrel{\text{def}}{=}-\sum p(X=x)\log p(X=x)$ denotes the binary entropy of X. $H(X|Y)\stackrel{\text{def}}{=}\sum_y p(y)H(X|Y=y)$ is the conditional entropy of X given Y.

section. The definitions follow those of [4, 7] (for communication complexity) and [3], with adaptations to the present model.

For arbitrary elements $a_k, \ldots, a_l$ we let $\langle a_k, \ldots, a_l \rangle$ or $\langle a_j \rangle_{j=k}^{l}$ denote the $(l+1-k)$-element sequence whose i'th element is a_{k+i-1}; for $k > l$ the sequence $\langle a_j \rangle_{j=k}^{l}$ is empty. A *bit* is an element of {0,1}. A *message*, or a *string*, is a (finite, possibly empty) sequence of bits. Let $\alpha_1, \ldots, \alpha_k$ and $\beta_1, \ldots, \beta_l$ be bits. The string $\langle \alpha_j \rangle_{j=1}^{k}$ is a *prefix* of the string $\langle \beta_j \rangle_{j=1}^{l}$ if $\alpha_j = \beta_j$ for $j = 1, \ldots, k$. It is a *proper prefix* if, in addition, $k < l$.

Let (X, Y) be a pair of random variables. The *support set* of (X, Y) is $S_{X,Y} = \{(x, y) : p(x, y) > 0\}$, and an *input* is an element of $S_{X,Y}$. A *protocol for* (X, Y) is a mapping ϕ that associates with each input (x, y) a finite sequence $\langle \sigma_1(x, y), \sigma_2(x, y), \ldots, \sigma_{m_\phi(x,y)}(x, y) \rangle$ of messages satisfying the three properties described below. Since $P_{\mathcal{X}}$ and $P_{\mathcal{Y}}$ alternate in transmitting these messages, all even-numbered messages are transmitted by one communicator and all odd-numbered messages by the other. We extend the messages indefinitely by letting $\sigma_i(x, y) \stackrel{\text{def}}{=} \Lambda$ (the empty string) for $i > m_\phi(x, y)$. This simplifies notation and proofs as we can assume that $\sigma_i(x, y)$ exists for all i. The following properties must hold for all $(x, y), (x, y'), (x', y) \in S_{X,Y}$ and all $i \geq 1$:

Separate-transmissions property

If message i is transmitted by $P_{\mathcal{X}}$ then $\langle \sigma_j(x, y) \rangle_{j=1}^{i-1} = \langle \sigma_j(x, y') \rangle_{j=1}^{i-1}$ implies that $\sigma_i(x, y) = \sigma_i(x, y')$,

and,

if message i is transmitted by $P_{\mathcal{Y}}$ then $\langle \sigma_j(x, y) \rangle_{j=1}^{i-1} = \langle \sigma_j(x', y) \rangle_{j=1}^{i-1}$ implies that $\sigma_i(x, y) = \sigma_i(x', y)$.

Implicit-termination property

If message i is transmitted by $P_{\mathcal{X}}$ then $\langle \sigma_j(x, y) \rangle_{j=1}^{i-1} = \langle \sigma_j(x', y) \rangle_{j=1}^{i-1}$ implies that neither $\sigma_i(x, y)$ nor $\sigma_i(x', y)$ is a *proper* prefix of the other,

and,

if message i is transmitted by $P_{\mathcal{Y}}$ then $\langle \sigma_j(x, y) \rangle_{j=1}^{i-1} = \langle \sigma_j(x, y') \rangle_{j=1}^{i-1}$ implies that neither $\sigma_i(x, y)$ nor $\sigma_i(x, y')$ is a *proper* prefix of the other.

Correct-decision property

$\langle \sigma_j(x, y) \rangle_{j=1}^{\infty} = \langle \sigma_j(x', y) \rangle_{j=1}^{\infty}$ implies that $x = x'$.

Several properties of protocols and complexity measures follow directly from these definitions and those given in the introduction. Some are described here. They are the present-model equivalent of results in [3, 8]. First we give an obvious lemma that has many ramifications.

Lemma 1 If (X_1, Y_1) and (X_2, Y_2) are two random pairs such that $S_{X_1,Y_1} \subseteq S_{X_2,Y_2}$, then $\hat{C}_m(X_1|Y_1) \le \hat{C}_m(X_2|Y_2)$ for all $m \in \{1,\ldots,\infty\}$.
Proof: Every protocol for (X_2, Y_2) is also a protocol for (X_1, Y_1). □

As an immediate consequence we have:

Corollary 1 If $S_{X_1,Y_1} = S_{X_2,Y_2}$ for two random pairs (X_1, Y_1) and (X_2, Y_2) then for all $m \in \{1,\ldots,\infty\}$,

$$\hat{C}_m(X_1|Y_1) = \hat{C}_m(X_2|Y_2) \ . \qquad \square$$

Two implications of this result are used throughout the paper:

1. $\hat{C}_m(X|Y)$ is a function only of $S_{X,Y}$, the support set of (X, Y); hence in determining $\hat{C}_m(X|Y)$ the precise values of p can be ignored.

2. There are only finitely many "essentially different" random pairs with values in a finite product-set; hence in Section 6 we can refer to "most."

These consequences were to be expected because of two assumptions made in formulating the problem. First, that $P_{\mathcal{Y}}$ cannot make any errors in determining X: however small $p(x, y)$ is, if $X = x$ and $Y = y$, then $P_{\mathcal{Y}}$ must eventually know that $X = x$. Second, that the complexity measure $\hat{C}_m(X|Y)$ is the number of bits exchanged in the worst-case (rather than on the average).

Other results can be obtained with little work. If $\sigma_k, \ldots, \sigma_l$ are strings, we let $[\sigma_k \ldots \sigma_l]$ or, equivalently, $[\sigma_j]_{j=k}^{l}$ denote the string obtained by their concatenation. Let ϕ be a protocol for (X, Y). The *codeword* that ϕ associates with the input (x, y) is the concatenation

$$\sigma_\phi(x, y) \stackrel{\text{def}}{=} [\sigma_j(x, y)]_{j=1}^{\infty}$$

of all messages transmitted under ϕ for that input. (Recall that ϕ is implicit in $\sigma_j(x, y)$.)

Theorem 1 Let ϕ be a protocol for a random pair (X, Y). If $(x', y), (x, y), (x, y') \in S_{X,Y}$ and one of $\sigma_\phi(x', y)$ and $\sigma_\phi(x, y')$ is a prefix of the other, then

$$\langle \sigma_j(x', y) \rangle_{j=1}^{\infty} = \langle \sigma_j(x, y) \rangle_{j=1}^{\infty} = \langle \sigma_j(x, y') \rangle_{j=1}^{\infty} \ .$$

Proof: We show by induction on i that $\langle\sigma_j(x',y)\rangle_{j=1}^{i} = \langle\sigma_j(x,y)\rangle_{j=1}^{i} = \langle\sigma_j(x,y')\rangle_{j=1}^{i}$. Induction hypothesis holds vacuously for $i=0$. Assume it holds for $i-1$, in particular $[\sigma_j(x',y)]_{j=1}^{i-1} = [\sigma_j(x,y')]_{j=1}^{i-1}$. Then one of $[\sigma_j(x',y)]_{j=i}^{\infty}$ and $[\sigma_j(x,y')]_{j=i}^{\infty}$ is a prefix of the other which implies that one of $\sigma_i(x',y)$ and $\sigma_i(x,y')$ is a prefix of the other. If the ith message is transmitted by $P_{\mathcal{X}}$ then, by the separate-transmissions property and the induction hypothesis, $\sigma_i(x,y) = \sigma_i(x,y')$, hence one of $\sigma_i(x,y)$ and $\sigma_i(x',y)$ is a prefix of the other. By the implicit-termination property, neither $\sigma_i(x,y)$ nor $\sigma_i(x',y)$ can be a proper prefix of the other, hence they must be the same and $\sigma_i(x',y) = \sigma_i(x,y) = \sigma_i(x,y')$. If the ith message is transmitted by $P_{\mathcal{Y}}$ then, symmetrically, $\sigma_i(x,y) = \sigma_i(x',y)$ by induction hypothesis and the separate-transmissions property, and, then, $\sigma_i(x,y) = \sigma_i(x,y')$ by the implicit-termination property, proving the induction step. □

If ϕ is a protocol for (X,Y), and (x,y), (x',y) are distinct inputs in $S_{X,Y}$, then, by the correct-decision property, $\langle\sigma_j(x,y)\rangle_{j=1}^{\infty} \neq \langle\sigma_j(x',y)\rangle_{j=1}^{\infty}$. The next corollary implies that not only the sequences of messages, but even the codewords themselves, must be different.

Corollary 2 Let ϕ be a transmission protocol for (X,Y). If (x',y), (x,y), $(x,y') \in S_{X,Y}$, and $x \neq x'$ then neither $\sigma_\phi(x',y)$ nor $\sigma_\phi(x,y')$ is a prefix of the other (in particular, they cannot be equal).

Proof: Assume that either $\sigma_\phi(x',y)$ or $\sigma_\phi(x,y')$ is a prefix of the other. From Theorem 1,

$$\langle\sigma_j(x',y)\rangle_{j=1}^{\infty} = \langle\sigma_j(x,y)\rangle_{j=1}^{\infty} = \langle\sigma_j(x,y')\rangle_{j=1}^{\infty} .$$

The first equality and the correct-decision property imply that $x = x'$. □

This enables us to apply traditional prefix-freeness results to our model. In Section 4 we will use it to apply Kraft's inequality; here, we prove a trivial yet useful lower bound on $\hat{C}_\infty(X|Y)$.

Equation (1) defined $P_{\mathcal{Y}}$'s ambiguity set $S_{X|Y}(y)$ to be the set of possible X values when $Y = y$. The last corollary implies that for all $y \in S_{\mathcal{Y}}$, the multiset[4] of codewords $\{\sigma_\phi(x,y) : x \in S_{X|Y}(y)\}$ is prefix free.

Corollary 3 For all (X,Y) pairs,

$$\hat{C}_\infty(X|Y) \geq \lceil \log \hat{\mu}_{X|Y} \rceil .$$

[4] A multiset allows multiplicity of elements. Hence, $\{1, 01, 01\}$ is prefix free as a set, but not as a multiset.

Proof: Let ϕ be a protocol for (X,Y). Corollary 2 implies that for all $y \in S_{\mathcal{Y}}$ the multiset $\{\sigma_\phi(x,y) : x \in S_{X|Y}(y)\}$ is prefix free. Hence, for some $x \in S_{X|Y}(y)$,

$$\hat{L}_\phi \geq l_\phi(x,y) \geq \lceil \log \mu_{X|Y}(y) \rceil \ .$$

This holds for all $y \in S_{\mathcal{Y}}$, proving the corollary. □

3 One-Way Complexity

$\hat{C}_1(X|Y)$, the one-way complexity of a random pair (X,Y), is the number of bits $P_{\mathcal{X}}$ must transmit in the worst case when $P_{\mathcal{Y}}$ is not permitted to transmit any feedback messages. Starting with $S_{X,Y}$, the support set of (X,Y), we define $G(X|Y)$, the *characteristic hypergraph* of (X,Y), and show that

$$\hat{C}_1(X|Y) = \lceil \log \mathcal{G} \rceil \ .$$

A hypergraph is an ordered pair (V,E) where V is a set, its elements called *vertices*, and E is a set of *hyperedges*: subsets of V. Two vertices are *adjacent* in a hypergraph if there is a hyperedge containing them both. A *coloring of a hypergraph*[5] $G = (V,E)$ is a mapping C defined over V such that $C(v_1) \neq C(v_2)$ whenever v_1 and v_2 are adjacent in G. The elements of the *range* $C(V) \stackrel{\text{def}}{=} \{C(v) : v \in V\}$ are *colors*. A *k-coloring of* G is a coloring of G whose range contains at most k colors. G is *k-colorable* if a k-coloring of G exists. The *chromatic number* $\chi(G)$ of G is the minimum k such that G is k-colorable.

Let (X,Y) be a random pair. For each y in $S_{\mathcal{Y}}$, the support set of Y, Equation (1) defined $S_{X|Y}(y)$ to be the set of possible x values when $Y = y$. The *characteristic hypergraph* $G(X|Y)$ of (X,Y) has $S_{\mathcal{X}}$ as its vertex set and the hyperedge $S_{X|Y}(y)$ for each $y \in S_{\mathcal{Y}}$.

Witsenhausen [1] defined a graph which is closely related to $G(X|Y)$ and linked its chromatic number to one-way communication. Let ϕ be a one-way protocol for a random pair (X,Y). The codeword $\sigma_\phi(x,y)$ of an input (x,y) consists of the message $\sigma_1(x,y)$. By the separate transmissions property, $\sigma_\phi(x,y)$ is independent of y and we denote it by $\sigma_\phi(x)$; it is the message transmitted by $P_{\mathcal{X}}$ when $X = x$. Define

$$\Sigma_\phi \stackrel{\text{def}}{=} \{\sigma_\phi(x) : x \in S_{\mathcal{X}}\}$$

[5]We denote both graphs and hypergraphs by G, reserving H for 'Hentropy'.

to be the set of all messages $P_{\mathcal{X}}$ may transmit under ϕ. If x_1 and x_2 are adjacent in $G(X|Y)$ then for some $y \in S_{\mathcal{Y}}$, both x_1 and x_2 are in $S_{X|Y}(y)$. By the correct-decision property, $\sigma_\phi(x_1) \neq \sigma_\phi(x_2)$ for $P_{\mathcal{Y}}$ to be able to deduce X. Hence, σ_ϕ is a coloring of $G(X|Y)$ and so

$$|\Sigma_\phi| \geq \mathcal{G} \ .$$

On the other hand, any coloring of $G(X|Y)$ suggests a one-way protocol where $P_{\mathcal{X}}$ transmits the color of X. Hence there exists a one-way protocol ϕ for (X, Y) with

$$|\Sigma_\phi| \leq \mathcal{G} \ .$$

Witsenhausen concluded that $\mathcal{G}$ is the minimum size of Σ_ϕ, i.e., the minimum number of possible messages transmitted by $P_{\mathcal{X}}$ in a one-way protocol. This proves that $\hat{C}_1(X|Y) \leq \lceil \log \mathcal{G} \rceil$. However, the set Σ_ϕ is not necessarily prefix free.[6] According to the implicit-termination property, messages have to be prefix free only for each given value of Y. Therefore, the above implies only that $\hat{C}_1(X|Y) \geq \lceil \log \mathcal{G} \rceil - 1$.

A string α *extends* another string β if β is a prefix of α. A string is *maximal* in a set if no other string in the set extends it. The next lemma proves that even the set of maximal codewords in Σ_ϕ has size of at least $\mathcal{G}$. This set is prefix free, hence at least one maximal codeword has length $\geq \lceil \log \mathcal{G} \rceil$.

Lemma 2 For all (X, Y) pairs,

$$\hat{C}_1(X|Y) = \lceil \log \mathcal{G} \rceil \ .$$

Proof: In view of the above, we prove only that $\hat{C}_1(X|Y) \geq \lceil \log \mathcal{G} \rceil$. Let ϕ be a protocol for (X, Y). For each $x \in S_{\mathcal{X}}$ let $\sigma'_\phi(x)$ be a maximal codeword in Σ_ϕ that extends $\sigma_\phi(x)$. The set

$$\Sigma'_\phi \stackrel{\text{def}}{=} \{\sigma'_\phi(x) : x \in S_{\mathcal{X}}\}$$

is prefix free and is contained in Σ_ϕ. If x_1 and x_2 are adjacent in $G(X|Y)$ then for some $y \in S_{\mathcal{Y}}$, both x_1 and x_2 are in $S_{X|Y}(y)$. Therefore

1. By the correct-decision property: $\sigma_\phi(x_1) \neq \sigma_\phi(x_2)$.
2. By the implicit-termination property: neither $\sigma_\phi(x_1)$ nor $\sigma_\phi(x_2)$ is a proper prefix of the other.

[6] Indeed, protocols with codewords that prefix each other can reduce the *average-case* complexity.

Therefore, neither $\sigma_\phi(x_1)$ nor $\sigma_\phi(x_2)$ is a (proper or improper) prefix of the other and so $\sigma'_\phi(x_1) \neq \sigma'_\phi(x_2)$. Hence σ'_ϕ too is a coloring of $G(X|Y)$ so Σ'_ϕ is a prefix-free set of cardinality $\geq \mathcal{G}$. Therefore, there must be an element in $\Sigma'_\phi \subseteq \Sigma_\phi$ whose length is at least $\lceil \log \mathcal{G} \rceil$. $\square$

This argument was implicitly used in Example 1 to show that $\hat{C}_1(X_t|Y_t) \geq \lceil \log t \rceil$. There, it was easy to see that all t messages have to be different and that no message could be a proper prefix of another. Both facts follow as the characteristic hypergraph in that case, $G(X_t|Y_t)$, is K_t, the complete graph on t vertices, whose chromatic number is t.

Unfortunately this characterization of $\hat{C}_1(X|Y)$ might not be of much practical help in determining $\hat{C}_1(X|Y)$. Given a random pair (X, Y), its characteristic hypergraph $G(X|Y)$ can be easily constructed but computing the chromatic number of a hypergraph is NP-complete, hence believed to be difficult.

4 The Limits of Interaction

Example 1 showed that for some (X, Y) pairs $\hat{C}_\infty(X|Y) \leq \log \hat{C}_1(X|Y) + 1$. Can $\hat{C}_\infty(X|Y)$ be any lower? are there (X, Y) pairs for which $\hat{C}_\infty(X|Y) \leq \log \hat{C}_1(X|Y)$? This section answers this question in the negative. A pair (X, Y) is *trivial* if $\hat{C}_1(X|Y) = 0$, that is, X is a deterministic function of Y. We show that for all nontrivial (X, Y) pairs,

$$\hat{C}_\infty(X|Y) \geq \lceil \log \hat{C}_1(X|Y) \rceil + 1 . \tag{3}$$

Therefore $\hat{C}_\infty(X|Y) = \lceil \log \hat{C}_1(X|Y) \rceil + 1$ is the largest reduction in communication achievable via feedback.

We divide this section into three subsections. In the first we prove a "bit" weaker result, namely that for all nontrivial (X, Y) pairs,

$$\hat{C}_\infty(X|Y) \geq \lceil \log \hat{C}_1(X|Y) \rceil .$$

Although very close to Inequality (3), this bound does not prove optimality of the two-message protocol of the league problem and of other protocols described in Section 5. Its proof however is simpler and more elegant than that of the stronger bound and is therefore given.

In subsection 4.2 we prove Inequality (3). The proof is recursive, relating the length of a message to the chromatic number of the characteristic hypergraph remaining after the message is sent. Though simple, the proof is rather long. With a bit of faith, the reader may choose to skip it at first.

Subsection 4.3 is very short. In it we show that the league problem is not unique in achieving the bound of Inequality (3). The bound is attained with equality by all (X, Y) pairs with maximum ambiguity two.

4.1 $\hat{C}_\infty(X|Y) \geq \lceil \log \hat{C}_1(X|Y) \rceil$

The proof uses simulation: given an arbitrary protocol for a pair (X, Y), we simulate it by a low-complexity one-way protocol. Simulation protocols have been used in communication complexity, e.g., Papdimitriou and Sipser [4]. There, however, all inputs are possible and interaction gain is obtained when the computed function is "irregular." Here, there is no function, and interaction helps only when the set of possible inputs is "irregular."

Let α and β be strings. We write $\beta \preceq \alpha$ ($\beta \not\preceq \alpha$) if β is (is not) a prefix of α and write $\beta \prec \alpha$ ($\beta \not\prec \alpha$) if β is (is not) a proper prefix of α.

Lemma 3 For all nontrivial (X, Y) pairs,

$$\hat{C}_\infty(X|Y) \geq \lceil \log \hat{C}_1(X|Y) \rceil \ .$$

Proof: Given a protocol ϕ for (X, Y) with worst-case complexity $\hat{L}_\phi$, we construct a one-way protocol whose worst-case complexity is $2^{\hat{L}_\phi}$. In Equation (1) we defined $S_{X|Y}(y)$ to be the set of possible X values when $Y = y$. We similarly define $P_\mathcal{X}$'s *ambiguity set* $S_{Y|X}(x)$ to be the set of possible Y values when $X = x$. Recall also that $\sigma_\phi(x, y)$ denotes the concatenation of all messages transmitted under ϕ for the input (x, y).

$P_\mathcal{X}$, given X, sequentially considers all $2^{\hat{L}_\phi}$ sequences of $\hat{L}_\phi$ bits. For each such sequence α, he transmits:

$$f_X(\alpha) \stackrel{\mathrm{def}}{=} \begin{cases} 1 & \text{if } \sigma_\phi(X, y) \preceq \alpha \text{ for some } y \in S_{Y|X}(X) \\ 0 & \text{otherwise.} \end{cases}$$

$P_\mathcal{Y}$ finds an $\hat{L}_\phi$-bit sequence α for which $f_X(\alpha) = 1$. He then finds an $x \in S_{X|Y}(Y)$ for which $\sigma_\phi(x, Y) \preceq \alpha$ and decides that $X = x$. To prove correctness of this protocol, we show:

1. There is an $\hat{L}_\phi$-bit sequence α for which $f_X(\alpha) = 1$ and $\sigma_\phi(X, Y) \preceq \alpha$.
2. If $x' \in S_{X|Y}(Y) - \{X\}$ then there is no $\hat{L}_\phi$-bit sequence α for which $f_X(\alpha) = 1$ and $\sigma_\phi(x', Y) \preceq \alpha$.

(1) holds because $(X,Y) \in S_{X,Y}$, hence $|\sigma_\phi(X,Y)| \le \hat{L}_\phi$. We prove (2) by showing that the existence of an α satisfying the two properties leads to a contradiction. If $f_X(\alpha) = 1$, then there is a $y' \in S_\mathcal{Y}$ such that $(X, y') \in S_{X,Y}$ and $\sigma_\phi(X, y') \preceq \alpha$. If, in addition, $\sigma_\phi(x', Y) \preceq \alpha$, then one of $\sigma_\phi(X, y')$ and $\sigma_\phi(x', Y)$ is a prefix of the other. By assumption, we also have $(X,Y) \in S_{X,Y}$ and $x' \neq X$, contradicting Corollary 2. □

4.2 $\hat{C}_\infty(X|Y) \ge \lceil \log \hat{C}_1(X|Y) \rceil + 1$

Lemma 2 related the one-way complexity of (X,Y) to the chromatic number of (X,Y)'s characteristic hypergraph: $\hat{C}_1(X|Y) = \lceil \log \mathcal{G} \rceil$. Since $\hat{C}_\infty(X|Y)$ is an integer, we need only show that $\hat{C}_\infty(X|Y) \ge \log\log \mathcal{G} + 1$. We do so by recursively considering the set of possible inputs after each message is transmitted.

Fix a random pair (X,Y) and a protocol ϕ for it. For $i \ge 0$ and arbitrary messages $\sigma_1, \ldots, \sigma_i$ let

$$S_{X,Y}(\langle \sigma_j \rangle_{j=1}^i) \stackrel{\text{def}}{=} \{(x,y) \in S_{X,Y} : \sigma_j(x,y) = \sigma_j \text{ for } j = 1, \ldots, i\}$$

be the set of inputs $(x,y) \in S_{X,Y}$ to which ϕ assigns messages $\sigma_1, \ldots, \sigma_i$. Additionally, let

$$S_\mathcal{X}(\langle \sigma_j \rangle_{j=1}^i) \stackrel{\text{def}}{=} \{x \in S_\mathcal{X} : (x,y) \in S_{X,Y}(\langle \sigma_j \rangle_{j=1}^i) \text{ for some } y \in S_\mathcal{Y}\}$$

be the set of x's such that for some y the protocol ϕ assigns $\sigma_1, \ldots, \sigma_i$ to the input (x,y), and define $S_\mathcal{Y}(\langle \sigma_j \rangle_{j=1}^i)$ similarly. It easy to verify that

$$S_{X,Y}(\langle \sigma_j \rangle_{j=1}^i) = S_{X,Y} \cap \left(S_\mathcal{X}(\langle \sigma_j \rangle_{j=1}^i) \times S_\mathcal{Y}(\langle \sigma_j \rangle_{j=1}^i) \right) .$$

Informally, if messages $\sigma_1, \ldots, \sigma_i$ are exchanged, then $P_\mathcal{X}$ and $P_\mathcal{Y}$ face a new communication problem where $P_\mathcal{X}$ knows a random variable $X \in S_\mathcal{X}(\langle \sigma_j \rangle_{j=1}^i)$, $P_\mathcal{Y}$ knows $Y \in S_\mathcal{Y}(\langle \sigma_j \rangle_{j=1}^i)$, and the support set of (X,Y) is $S_{X,Y}(\langle \sigma_j \rangle_{j=1}^i)$. If $Y = y$ then, as far as $P_\mathcal{Y}$ knows, X can be any element of the ambiguity set $\{x : (x,y) \in S_{X,Y}(\langle \sigma_j \rangle_{j=1}^i)\}$. Therefore, communication can stop after messages $\sigma_1, \ldots, \sigma_i$ only if for each $y \in S_\mathcal{Y}(\langle \sigma_j \rangle_{j=1}^i)$ there is exactly one $x \in S_\mathcal{X}(\langle \sigma_j \rangle_{j=1}^i)$ such that $(x,y) \in S_{X,Y}(\langle \sigma_j \rangle_{j=1}^i)$.

To derive bounds using this property we further pursue the analogy to hypergraph coloring. The messages $\sigma_1, \ldots, \sigma_i$ are *possible* if $S_{X,Y}(\langle \sigma_j \rangle_{j=1}^i) \neq \emptyset$, namely, they occur for some input. For $i \ge 0$ and a sequence $\sigma_1, \ldots, \sigma_i$ of possible messages, let $G(\langle \sigma_j \rangle_{j=1}^i)$ be the characteristic hypergraph of any (X,Y) pair whose support set is $S_{X,Y}(\langle \sigma_j \rangle_{j=1}^i)$. Its vertex set is $S_\mathcal{X}(\langle \sigma_j \rangle_{j=1}^i)$ and each $y \in S_\mathcal{Y}(\langle \sigma_j \rangle_{j=1}^i)$ contributes the hyperedge $\{x : (x,y) \in S_{X,Y}(\langle \sigma_j \rangle_{j=1}^i)\}$.

From the last paragraph, communication can stop after messages $\sigma_1,\ldots,\sigma_i$ only if each hyperedge of $G(\langle\sigma_j\rangle_{j=1}^i)$ contains at most one element, i.e., only if $\chi(G(\langle\sigma_j\rangle_{j=1}^i))=1$.

Roughly speaking,[7] Lemma 5 shows that for every sequence $\sigma_1,\ldots,\sigma_i$ of possible messages, there is a message σ_{i+1} such that

$$\log\chi(G(\langle\sigma_j\rangle_{j=1}^{i+1}))\geq\log\chi(G(\langle\sigma_j\rangle_{j=1}^{i}))\cdot 2^{-|\sigma_{i+1}|}\ .$$

Recursively then, for all i there exists a possible sequence $\sigma_1,\ldots,\sigma_i$ of messages such that

$$\log\chi(G(\langle\sigma_j\rangle_{j=1}^{i}))\geq\log\mathcal{G}\cdot 2^{-(|\sigma_1|+\cdots+|\sigma_i|)}\ .$$

Let m be any integer. For every possible sequence $\sigma_1,\ldots,\sigma_i$,

$$|\sigma_1|+\cdots+|\sigma_m|\leq\hat{L}_\phi$$

and, if ϕ is m-message then

$$\chi(G(\langle\sigma_j\rangle_{j=1}^{m}))=1\ .$$

Therefore, for at least one sequence $\sigma_1,\ldots,\sigma_m$,

$$1\approx\log\chi(G(\langle\sigma_j\rangle_{j=1}^{m}))\geq\log\mathcal{G}\cdot 2^{-(|\sigma_1|+\cdots+|\sigma_m|)}\geq\log\mathcal{G}\cdot 2^{-\hat{L}_\phi}\ .$$

This implies $\hat{L}_\phi\geq\log\log\mathcal{G}$ for every m-message protocol ϕ hence $\hat{C}_m(X|Y)\geq\lceil\log\hat{C}_1(X|Y)\rceil$. A more careful analysis, carried out in the proof of Theorem 2 yields the exact lower bound: $\hat{C}_\infty(X|Y)\geq\lceil\log\hat{C}_1(X|Y)\rceil+1$.

We begin with some results concerning hypergraphs. Let $G=(V,E)$ be a hypergraph, and let $V'\subseteq V$ and $E'\subseteq E$. The *vertex restriction* $G|_{V'}$ of G to V' is the hypergraph whose vertex set is V' and whose edge set is $\{e\cap V' : e\in E\}$. The *edge restriction* $G/_{E'}$ of G to E' is the hypergraph whose edge set is E' and whose vertex set is $\{v : v\in e' \text{ for some } e'\in E'\}$.

When $P_\mathcal{X}$ transmits a message he restricts the set of possible X values; similarly for messages transmitted by $P_\mathcal{Y}$. It is easy to verify that if message $\sigma_1,\ldots,\sigma_i$ have been transmitted then:

1. If σ_{i+1} is sent by $P_\mathcal{X}$ then $G(\langle\sigma_j\rangle_{j=1}^{i+1})=G(\langle\sigma_j\rangle_{j=1}^{i})|_{S_\mathcal{X}(\langle\sigma_j\rangle_{j=1}^{i+1})}$.
2. If σ_{i+1} is sent by $P_\mathcal{Y}$ then $G(\langle\sigma_j\rangle_{j=1}^{i+1})=G(\langle\sigma_j\rangle_{j=1}^{i})/_{S_\mathcal{Y}(\langle\sigma_j\rangle_{j=1}^{i+1})}$.

[7]The precise statements are given in the appropriate lemmas and theorem.

This relates vertex and edge restrictions to messages sent by $P_{\mathcal{X}}$ and $P_{\mathcal{Y}}$. The next Lemma implies that if a message can assume only a few possible values then at least one of the resulting restricted hypergraphs has a large chromatic number.

Lemma 4 Let $G = (V, E)$ be a hypergraph with chromatic number $\chi(G)$.

1. If $V_1, \ldots, V_k$ partition V then $\chi(G|_{V_1}) + \cdots + \chi(G|_{V_k}) \geq \chi(G)$.
2. If $E_1, \ldots, E_k$ partition E then $\chi(G/_{E_1}) \cdot \ldots \cdot \chi(G/_{E_k}) \geq \chi(G)$.

Proof:

1. Let C_i be a $\chi(G|_{V_i})$-coloring of $G|_{V_i}$. Define a mapping C over V as follows: $C(v) = (i, C_i(v))$ where i is the unique integer such that $v \in V_i$. C is a valid coloring of G because if v', v'' are in different V_i's then the first components of $C(v')$ and $C(v'')$ differ whereas if v' and v'' are neighbors in the same V_i, then the second components of $C(v')$ and $C(v'')$ differ. The number of colors used is $|C(V)| = \sum_{i=1}^{k} |C(V_i)| = \sum_{i=1}^{k} \chi(G|_{V_i})$.
2. Let C_i be a $\chi(G/_{E_i})$-coloring of $G/_{E_i}$. Define $C(v) = (C_1(v), \ldots, C_k(v))$. Then $\{v', v''\} \subseteq e$ for some $e \in E$ implies $\{v', v''\} \subseteq e \in E_i$ for some $1 \leq i \leq k$, hence the ith components of $C(v')$ and $C(v'')$ differ. $\square$

Assume for example that after $\sigma_1, \ldots, \sigma_i$ were transmitted $P_{\mathcal{X}}$ transmits a k-bit message. Each possible message corresponds to a vertex restriction of $G(\langle\sigma_j\rangle_{j=1}^{i})$ and at least one message, σ_{i+1}, corresponds to a hypergraph $G(\langle\sigma_j\rangle_{j=1}^{i+1})$ with $\chi(G(\langle\sigma_j\rangle_{j=1}^{i+1})) \geq \chi(G(\langle\sigma_j\rangle_{j=1}^{i})) \cdot 2^{-k}$. This intuitive argument has two possible pitfalls. First, $P_{\mathcal{X}}$ can transmit messages of different lengths and assign longer messages to restricted hypergraphs with smaller chromatic number. More intricate is the possibility of a message being a prefix of another when no confusion is caused. This could lead to more efficient transmissions as k bits can correspond to 2^{k+1} (rather than 2^k) different messages. The next lemma handles both cases. A hypergraph is *trivial* if it contains no vertices or no edges.

Lemma 5 Let $G = (V, E)$ be a nontrivial hypergraph. Then

1. If $\psi : V \to \{0,1\}^*$ is such that $\psi(v) \not\prec \psi(v')$ (proper) whenever v and v' are adjacent then there is a $\sigma_0 \in \psi(V)$ such that $\chi(G|_{\psi^{-1}(\sigma_0)}) \geq \chi(G) \cdot 2^{-|\sigma_0|}$.
2. If $\psi : E \to \{0,1\}^*$ is such that $\psi(e) \not\prec \psi(e')$ (proper) whenever e and e' share a vertex then there is a $\sigma_0 \in \psi(E)$ such that $\log(\chi(G/_{\psi^{-1}(\sigma_0)})) \geq \log(\chi(G)) \cdot 2^{-|\sigma_0|}$. $\square$

Rather than prove each part separately, we prove a more general result from which both parts follow easily.

Lemma 6 Let S be a nonempty set, $\psi : S \to \{0,1\}^*$ a mapping that assigns a string to each element of S, and $\rho : 2^S \to [0, \infty)$ a mapping that assigns a nonnegative real value to each subset of S such that:

1. $\rho(A \bigcup B) \leq \rho(A) + \rho(B)$ for all $A, B \subseteq S$.

2. $\rho(A \bigcup B) = \max\{\rho(A), \rho(B)\}$ for all $A, B \subseteq S$ such that $\psi(a) \prec \psi(b)$ for all $a \in A$ and $b \in B$.

Then, there is a string σ_0 such that $\rho(\psi^{-1}(\sigma_0)) \geq \rho(S) \cdot 2^{-|\sigma_0|}$.

Proof: Consider the range $\psi(S)$ of ψ. As a set of messages, it is either prefix free or not.

1. In case $\psi(S)$ is prefix free. The first assumption of the lemma implies

$$\rho(S) = \rho(\bigcup_{\sigma \in \psi(S)} \psi^{-1}(\sigma)) \leq \sum_{\sigma \in \psi(S)} \rho(\psi^{-1}(\sigma)) \ .$$

   Therefore, by Kraft's inequality,

$$\sum_{\sigma \in \psi(S)} 2^{-|\sigma|} \leq 1 \leq \sum_{\sigma \in \psi(S)} \frac{\rho(\psi^{-1}(\sigma))}{\rho(S)} \ .$$

   Hence there must be a message $\sigma_0 \in \psi(S)$ such that $\rho(\psi^{-1}(\sigma_0)) \geq \rho(S) \cdot 2^{-|\sigma_0|}$.

2. In case $\psi(S)$ is not prefix free. For a set Σ of messages, let its *prefix number*, $p.n.(\Sigma) \stackrel{\text{def}}{=} |\{\sigma \in \Sigma : \sigma \prec \sigma' \text{ for some } \sigma' \in \Sigma\}|$, be the number of non-maximal messages in Σ. We show that if the lemma does not hold for a mapping ψ with a positive $p.n.(\psi(S))$, then it does not hold for a mapping ψ' with $p.n.(\psi'(S)) < p.n.(\psi(S))$. This implies that the lemma does not hold for a mapping with a prefix-free range, contradicting the first part of the proof.

   Let ψ be a mapping satisfying the conditions of the lemma but not its consequence, that is, $\rho(\psi^{-1}(\sigma)) < \rho(S) \cdot 2^{-|\sigma|}$ for all $\sigma \in \psi(S)$. Consider the forest whose vertex set is the range $\psi(S)$ of ψ and where message $\sigma'' \in \psi(S)$ is a child of another message $\sigma \in \psi(S)$ if $\sigma \prec \sigma''$ and there is no $\sigma' \in \psi(S)$ such that $\sigma \prec \sigma' \prec \sigma''$ (all prefixes proper). The forest is not necessarily binary, and it is a tree if and only if there is a message that is a prefix of all other messages.

Since $\psi(S)$ is not prefix free there is a tree in the forest which is not a leaf. This tree contains an internal node all of whose descendents are leaves. Let σ_* be that node and let $\sigma_1, \ldots, \sigma_k$ be its descendents.

Define the new mapping ψ' as follows:

$$\psi'(s) = \begin{cases} \sigma_* & \text{if } \psi(s) = \sigma_i \text{ for some } 1 \leq i \leq k \\ \psi(s) & \text{otherwise} \end{cases}$$

The first requirement of the lemma clearly holds because ρ was not changed. The second requirement is satisfied because $\psi'(a) \prec \psi'(b)$ implies $\psi(a) \prec \psi(b)$. The consequence of the lemma does not hold because:

$$\begin{aligned}
\rho(\psi'^{-1}(\sigma_*)) &= \rho(\psi^{-1}(\sigma_*) \bigcup \psi^{-1}(\sigma_1) \bigcup \cdots \bigcup \psi^{-1}(\sigma_k)) \\
&= \max\{\rho(\psi^{-1}(\sigma_*)), \rho(\psi^{-1}(\sigma_1) \bigcup \cdots \bigcup \psi^{-1}(\sigma_k))\} \\
&\leq \max\{\rho(\psi^{-1}(\sigma_*)), \sum_{i=1}^{k} \rho(\psi^{-1}(\sigma_i))\} \\
&< \rho(S) \cdot \max\{2^{-|\sigma_*|}, \sum_{i=1}^{k} 2^{-|\sigma_i|}\} \\
&= \rho(S) \cdot \max\{2^{-|\sigma_*|}, 2^{-|\sigma_*|} \sum_{i=1}^{k} 2^{-(|\sigma_i| - |\sigma_*|)}\} \\
&\leq \rho(S) \cdot \max\{2^{-|\sigma_*|}, 2^{-|\sigma_*|}\} \\
&= \rho(S) \cdot 2^{-|\sigma_*|}
\end{aligned}$$

where we have used the fact that none of the σ_i's is a prefix of any other hence $\sum_{i=1}^{k} 2^{-(|\sigma_i| - |\sigma_*|)} \leq 1$. □

The proof of Lemma 5 now follows by a proper choice of ρ.

Part 1: Define $\rho : 2^V \rightarrow Z^+$ by $\rho(A) \stackrel{\text{def}}{=} \chi(G|_A)$ for all $A \subseteq V$. Part 1 of Lemma 4 provides the first requirement of the lemma. The second requirement is satisfied because $\psi(a) \prec \psi(b)$ for all $a \in A$ and $b \in B$ implies that no element in A is adjacent to any element in B. Therefore the same colors can be used for vertices of A and B.

Part 2: Define $\rho : 2^E \rightarrow R^+$ by $\rho(A) \stackrel{\text{def}}{=} \log \chi(G/_A)$ for all $A \subseteq E$. Part 2 of Lemma 4 implies $\rho(A \bigcup B) = \log(\chi(G/_{A \bigcup B})) \leq \log(\chi(G/_A) \cdot \chi(G/_B)) = \log \chi(G/_A) + \log \chi(G/_B) =$

$\rho(A)+\rho(B)$. The second requirement is satisfied because if $\psi(a) \prec \psi(b)$ for all $a \in A$ and $b \in B$ then no edge in A shares a vertex with an edge in B. Hence $G/_A$ and $G/_B$ are disjoint and the same colors can be used for both.

We combine the results to prove the lower bound for $\hat{C}_\infty(X|Y)$:

Theorem 2 For all nontrivial (X,Y) pairs

$$\hat{C}_\infty(X|Y) \geq \lceil \log \hat{C}_1(X|Y) \rceil + 1 \ .$$

Proof: Let ϕ be a protocol for (X,Y). We show by induction that for all $i \geq 0$ there is a possible sequence $\sigma_1,\ldots,\sigma_i$ of messages such that either

A. $\chi(G(\langle\sigma_j\rangle_{j=1}^{i})) \geq 2$ and $\sum\limits_{j=1}^{i} |\sigma_j| \geq \log\log \mathcal{G} - \log\log \chi(G(\langle\sigma_j\rangle_{j=1}^{i}))$.

or

B. $\chi(G(\langle\sigma_j\rangle_{j=1}^{i})) = 1$ and $\sum\limits_{j=1}^{i} |\sigma_j| \geq \log\log \mathcal{G} + 1$.

Induction basis: Assertion (A) holds for $i=0$ as nontrivial (X,Y) pairs have $\mathcal{G} \geq 2$ and both sides of the second inequality are 0. Induction step: If (B) holds for $\sigma_1,\ldots,\sigma_i$ then (B) holds for $\sigma_1,\ldots,\sigma_{i+1}$ where σ_{i+1} is any message possible under ϕ. If (A) holds for $\sigma_1,\ldots,\sigma_i$ then there are two cases:

1. $P_\mathcal{Y}$ transmits message $i+1$. From Lemma 5 there must be a message σ_{i+1} such that

   $$\log \chi(G(\langle\sigma_j\rangle_{j=1}^{i+1})) \geq \log \chi(G(\langle\sigma_j\rangle_{j=1}^{i})) \cdot 2^{-|\sigma_{i+1}|} \ .$$

   Since (A) holds for $\sigma_1,\ldots,\sigma_i$ both sides of the inequality are strictly positive. Hence $\chi(G(\langle\sigma_j\rangle_{j=1}^{i+1})) \geq 2$ and $|\sigma_{i+1}| \geq \log\log \chi(G(\langle\sigma_j\rangle_{j=1}^{i})) - \log\log \chi(G(\langle\sigma_j\rangle_{j=1}^{i+1}))$.
   Combined with the induction hypothesis, we see that (A) holds again:
   $\sum\limits_{j=1}^{i+1} |\sigma_j| \geq \log\log \mathcal{G} - \log\log \chi(G(\langle\sigma_j\rangle_{j=1}^{i+1}))$.

2. $P_\mathcal{X}$ transmits message $i+1$. From Lemma 5 there must be a message σ_{i+1} such that

   $$\chi(G(\langle\sigma_j\rangle_{j=1}^{i+1})) \geq \chi(G(\langle\sigma_j\rangle_{j=1}^{i})) \cdot 2^{-|\sigma_{i+1}|} \ .$$

   If $\chi(G(\langle\sigma_j\rangle_{j=1}^{i+1})) = 1$ then $|\sigma_{i+1}| \geq \lceil \log \chi(G(\langle\sigma_j\rangle_{j=1}^{i}))\rceil \geq \log\log \chi(G(\langle\sigma_j\rangle_{j=1}^{i})) + 1$. Hence $\sum_{j=1}^{i+1} |\sigma_j| \geq \log\log \mathcal{G} + 1$ and (B) holds.

   Otherwise $\chi(G(\langle\sigma_j\rangle_{j=1}^{i+1})) \geq 2$. Using simple monotonicity properties of the function $x - \log x$ we obtain

   $$\begin{aligned} |\sigma_{i+1}| &\geq \log \chi(G(\langle\sigma_j\rangle_{j=1}^{i})) - \log \chi(G(\langle\sigma_j\rangle_{j=1}^{i+1})) \\ &\geq \log\log \chi(G(\langle\sigma_j\rangle_{j=1}^{i})) - \log\log \chi(G(\langle\sigma_j\rangle_{j=1}^{i+1})) \end{aligned}$$

for all (integer) values of $\chi(G(\langle\sigma_j\rangle_{j=1}^{i})) \geq \chi(G(\langle\sigma_j\rangle_{j=1}^{i+1})) \geq 2$ except $\chi(G(\langle\sigma_j\rangle_{j=1}^{i})) = 3$ and $\chi(G(\langle\sigma_j\rangle_{j=1}^{i+1})) = 2$. For this case we verify directly that $|\sigma_{i+1}| \geq 1 \geq \log\log 3 - \log\log 2$. As before we can combine with the induction hypothesis to see that (A) holds again.

To complete the proof, let ϕ be an m-message protocol for (X, Y). All possible message sequences $\sigma_1, \ldots, \sigma_m$ have $\chi(G(\langle\sigma_j\rangle_{j=1}^{m})) = 1$. Hence (B) must hold for some sequence $\sigma_1, \ldots, \sigma_m$. Let (x, y) be any element of $S_{X,Y}(\langle\sigma_j\rangle_{j=1}^{m})$. Then

$$\hat{L}_\phi \geq l_\phi(x, y) = \sum_{j=1}^{m} |\sigma_j| \geq \log\log \mathcal{G} + 1 \ .$$

Therefore, for any finite m,

$$\hat{C}_m(X|Y) \geq \lceil \log\log \mathcal{G} \rceil + 1 = \lceil \log \hat{C}_1(X|Y) \rceil + 1 \ ,$$

and the theorem follows. □

We reap the benefits of the tedious work by applying Theorem 2 to precisely determine $\hat{C}_m(X_t|Y_t)$ for all m in the league problem of Example 1.

Example 2 It is now easy to see that $\lceil \log\log t \rceil + 1$ bits are necessary for the league problem with two or even more messages. Example 1 showed that $\hat{C}_1(X_t|Y_t) = \lceil \log t \rceil$, and that $\hat{C}_2(X_t|Y_t) \leq \lceil \log\log t \rceil + 1$. Theorem 2 implies that $\hat{C}_\infty(X_t|Y_t) \geq \lceil \log \hat{C}_1(X_t|Y_t) \rceil + 1 \geq \lceil \log\log t \rceil + 1$. Hence $\hat{C}_2(X_t|Y_t) = \cdots = \hat{C}_\infty(X_t|Y_t) = \lceil \log\log t \rceil + 1$. □

4.3 If $\hat{\mu}_{X|Y} = 2$ then $\hat{C}_\infty(X|Y) = \lceil \log \hat{C}_1(X|Y) \rceil + 1$

Equation (2) defined $\hat{\mu}_{X|Y}$, the maximum ambiguity of (X, Y), as the maximum number of X values possible with any given Y value. In this subsection we show that for all (X, Y) pairs with maximum ambiguity two, the bound of Inequality (3) is tight. It is therefore the tightest general bound on $\hat{C}_\infty(X|Y)$ that can be expressed in terms of m-message complexities.

Theorem 3 For all (X, Y) pairs with $\hat{\mu}_{X|Y} = 2$,

$$\hat{C}_\infty(X|Y) = \lceil \log \hat{C}_1(X|Y) \rceil + 1 \ .$$

Proof: We describe a 2-message protocol whose worst-case complexity is $\lceil \log \hat{C}_1(X|Y) \rceil + 1$. This proves $\hat{C}_\infty(X|Y) \leq \lceil \log \hat{C}_1(X|Y) \rceil + 1$. Theorem 2 provides the converse inequality.

The protocol is essentially that of the league problem, but it uses colors instead of actual X-values. $P_{\mathcal{X}}$ and $P_{\mathcal{Y}}$ agree on a $\mathcal{G}$-coloring of $G(X|Y)$ and on a $\lceil\log\mathcal{G}\rceil$-bit encoding of each color. $P_{\mathcal{Y}}$ knows Y and therefore knows the two possible values of X and their (necessarily different) colors. He finds a bit location in which the encodings of the two corresponding colors differ and transmits that location using $\lceil\log\lceil\log\mathcal{G}\rceil\rceil$ bits. Then, $P_{\mathcal{X}}$ responds with one bit describing the value of that bit in the encoding of the color of X. The total number of transmitted bits is $\lceil\log\lceil\log\mathcal{G}\rceil\rceil + 1$ and Lemma 2 said that $\hat{C}_1(X|Y) = \lceil\log\mathcal{G}\rceil$. □

Example 1 illustrated a special case. For each Y value (a match) there were two possible X values (winning teams) hence $\hat{\mu}_{X|Y}$ was 2. This was a simple case to analyze as all 2-subsets of $S_{\mathcal{X}}$ could occur. In general, $S_{X,Y}$ is much less regular.

Example 3 Modifying Example 1, assume an infinite league of teams named $1, 2, \ldots$. Two random teams I and J play each other and team $K \in \{I, J\}$ wins. $P_{\mathcal{Y}}$ knows who played, $P_{\mathcal{X}}$ knows who won. This time, however, teams I and J are always picked so that $|I - J| \leq \ell$. (Perhaps the names correspond to locations, or strengths, and two teams too far in distance or strength cannot play each other).

The vertices $1, \ldots, \ell$ form a complete subgraph of the infinite graph $G(X|Y)$, hence $\mathcal{G} \geq \ell$. On the other hand, the mapping $C(i) = i \bmod \ell$ is an ℓ-coloring of $G(X|Y)$. Hence $\mathcal{G} = \ell$ and Theorem 3 implies that $\hat{C}_1(X|Y) = \lceil\log\ell\rceil$ and $\hat{C}_\infty(X|Y) = \lceil\log\log\ell\rceil + 1$.

One optimal one-way protocol instructs $P_{\mathcal{X}}$ to transmit the $\lceil\log\ell\rceil$-bit binary representation of $K \bmod \ell$. In one optimal two-message protocol, $P_{\mathcal{Y}}$ uses $\lceil\log\log\ell\rceil$ bits to transmit a location where the $\lceil\log\ell\rceil$-bit binary representation of $I \bmod \ell$ differs from that of $J \bmod \ell$, and $P_{\mathcal{X}}$ transmits the value of this bit for the winning team. □

5 Two Messages are Almost Optimal

In the last section we saw that for all (X,Y) pairs $\hat{C}_\infty(X|Y) \geq \lceil\log\hat{C}_1(X|Y)\rceil + 1$, and that for some pairs equality is achievable: $\hat{C}_1(X|Y) = \frac{1}{2}2^{\hat{C}_\infty(X|Y)}$. In a sense, this is a negative result implying that $\hat{C}_1(X|Y)$, the number of bits required when no feedback is permitted, may be exponentially higher than $\hat{C}_\infty(X|Y)$, the minimum number achievable when no restrictions apply. In this section we prove that $\hat{C}_\infty(X|Y)$ can almost be achieved with just one feedback message: $P_{\mathcal{Y}}$ transmits one message to $P_{\mathcal{X}}$, then $P_{\mathcal{X}}$ responds with one message. Precisely:

$$\hat{C}_2(X|Y) \leq 4\hat{C}_\infty(X|Y) + 3$$

for all (X, Y) pairs. To prove this bound we present a method that, given any random pair (X, Y), constructs a two-message protocol whose complexity is close to $\hat{C}_\infty(X|Y)$.

This sharply contrasts with communication complexity. Papadimitriou and Sipser [4] showed a function for which one-message protocols require $\sqrt{t}/(2\log t)$ transmitted bits while there are two-message protocols for that function requiring only $2\log t + 1$ bits. Duris, Galil, and Schnitger [5] showed that for every m, there is a function whose m-message complexity is $\Omega(\sqrt{t}/(m^4 \log t))$ but its $(m+1)$-message complexity is only $O(m \log t)$. Hence, in communication complexity there can be an almost-exponential gap between m- and $(m+1)$-message complexities. Here, all complexity measures between $\hat{C}_2(X|Y)$ and $\hat{C}_\infty(X|Y)$ are at most a factor of four apart and only $\hat{C}_1(X|Y)$ can be exponentially higher.

Equation (2) defined the maximum ambiguity $\hat{\mu}_{X|Y}$ of a pair (X, Y) to be the maximum number of X values possible with any Y value. Corollary 3 showed that $\lceil \log \hat{\mu}_{X|Y} \rceil$ is a trivial lower bound on $\hat{C}_\infty(X|Y)$. The results proven here show that within a factor of four, $\log \hat{\mu}_{X|Y}$ is a reasonable estimate for $\hat{C}_m(X|Y)$ for all $m \geq 2$.

Lemma 2 showed that $\hat{C}_1(X|Y) = \lceil \log \mathcal{G} \rceil$ for all (X, Y) pairs. With one exception, it is very difficult to precisely determine $\hat{C}_m(X|Y)$ for large m. The exception is when the maximum ambiguity of (X, Y) is two. Whenever $\hat{\mu}_{X|Y} = 2$ we can completely characterize $\hat{C}_m(X|Y)$ for all m.

We therefore divide this section into two subsections corresponding to the value of $\hat{\mu}_{X|Y}$. In Subsection 5.1 we show that all (X, Y) pairs with $\hat{\mu}_{X|Y} = 2$,

$$\hat{C}_2(X|Y) = \cdots = \hat{C}_\infty(X|Y) = \lceil \log \hat{C}_1(X|Y) \rceil + 1 \ .$$

In Subsection 5.2 we show that for all (X, Y) pairs,

$$\hat{C}_2(X|Y) \leq 4\hat{C}_\infty(X|Y) + 3 \ ,$$

even if $\hat{\mu}_{X|Y} > 2$.

5.1 $\hat{\mu}_{X|Y} = 2$: One Feedback Message is Optimal

We prove that when the maximum ambiguity of a pair (X, Y) is two, two messages suffice to achieve minimum communication.

Corollary 4 For all (X, Y) pairs with $\hat{\mu}_{X|Y} \leq 2$,

$$\hat{C}_2(X|Y) = \hat{C}_3(X|Y) = \ldots = \hat{C}_\infty(X|Y) \ . \qquad \square$$

In fact, this result was already proven in Theorem 3. The theorem's proof also showed that all (X,Y) pairs with $\hat{\mu}_{X|Y}=2$ have an optimal protocol where the informer, $P_{\mathcal{X}}$, transmits only one bit and the recipient, $P_{\mathcal{Y}}$, transmits the rest. However, the proof was geared toward achievability: showing that $\hat{C}_2(X|Y)$ is small. While it provides a good introduction to the protocols of the next subsection, it sheds little if any light on the reason why two messages may not achieve $\hat{C}_\infty(X|Y)$ when $\hat{\mu}_{X|Y}>2$.

We therefore present another proof that relies upon the characteristic-hypergraph structure of the problem. It provides more insight into the advantages of feedback messages and explains why two messages suffice when $\hat{\mu}_{X|Y}=2$ but do not necessarily suffice for larger $\hat{\mu}_{X|Y}$.

A *graph* is a hypergraph where each hyperedge contains at most 2 vertices. When $\hat{\mu}_{X|Y}=2$, the characteristic hypergraph of (X,Y) is a graph. For graphs, the converse of Lemma 4 holds. It is this property of $G(X|Y)$ that enables us to determine $\hat{C}_2(X|Y)$ and achieve it with just one feedback message.

Lemma 7 Let $G=(V,E)$ be a graph with chromatic number $\chi(G)$.

1. If $\chi_1+\ldots+\chi_k\geq\chi(G)$ then there exists a partition $V_1,\ldots,V_k$ of V such that $G|_{V_i}$ is χ_i-colorable for $i=1,\ldots,k$. (This part is true for hypergraphs in general.)

2. If $\chi_1\cdot\ldots\cdot\chi_k\geq\chi(G)$ then there exists a partition $E_1,\ldots,E_k$ of E such that $G/_{E_i}$ is χ_i-colorable for $i=1,\ldots,k$.

Proof:

1. Let $C:V\to\{1,\ldots,\chi(G)\}$ be a $\chi(G)$-coloring of G. For $i=1,\ldots,k$ define
$$V_i\stackrel{\text{def}}{=}\{v\in V:\sum_{j=1}^{i-1}\chi_j<C(v)\leq\sum_{j=1}^{i}\chi_j\}\ .$$
Since $\sum_{j=1}^{k}\chi_i\geq\chi(G)$, the sets $V_1,\ldots,V_k$ partition V. Also, for all $i=1,\ldots,k$, the restriction of C to V_i is a valid χ_i-coloring of $G|_{V_i}$.

2. The lemma holds for $k=1$. Assuming it holds for $k-1$, we prove it holds for k. Define $\chi\stackrel{\text{def}}{=}\prod_{j=1}^{k-1}\chi_j$. Then $\chi\cdot\chi_k\geq\chi(G)$. Let $C:V\to\{1,\ldots,\chi(G)\}$ be a $\chi(G)$-coloring of G. For $i=1,\ldots,\chi_k$ define
$$V_i\stackrel{\text{def}}{=}\{v\in V:(i-1)\cdot\chi<C(v)\leq i\cdot\chi\}\ .$$

Since $\chi \cdot \chi_k \geq \chi(G)$, the V_i's partition V. Let E_s be the set of edges that connect two vertices in V_i for some i and let E_k be the set of edges that connect vertices in distinct V_i's. For $i = 1, \ldots, \chi_k$, the graph $(G/_{E_k})|_{V_i}$ is 1-colorable, therefore $\chi(G/_{E_k}) \leq \chi_k$. For $i = 1, \ldots, \chi_k$, the graph $(G/_{E_s})|_{V_i}$ is χ-colorable and these graphs are all disjoint, therefore $\chi(G/_{E_s}) \leq \chi = \prod_{i=1}^{k-1} \chi_i$. By induction hypothesis, there is a partition $E_1, \ldots, E_{k-1}$ of E_s such that $\chi(G/_{E_i}) = \chi((G/_{E_s})/_{E_i}) \leq \chi_i$. Since E_s and E_k partition E, so do $E_1, \ldots, E_k$. □

Together with Lemma 4 this lemma implies that if $G = (V, E)$ is a graph with chromatic number $\chi(G)$ and $\chi_1, \ldots, \chi_k$ are integers then

1. There is a partition $V_1, \ldots, V_k$ of V such that $G|_{V_i}$ is χ_i-colorable for $i = 1, \ldots, k$ if and only if $\chi_1 + \ldots + \chi_k \geq \chi(G)$.

2. There is a partition $E_1, \ldots, E_k$ of E such that $G/_{E_i}$ is χ_i-colorable for $i = 1, \ldots, k$ if and only if $\chi_1 \cdot \ldots \cdot \chi_k \geq \chi(G)$.

The lemma suggests that bits transmitted by $P_{\mathcal{Y}}$ are "more powerful" than those transmitted by $P_{\mathcal{X}}$. Whereas each bit transmitted by $P_{\mathcal{X}}$ can only reduce the chromatic number of the characteristic hypergraph to half of its previous value, each bit transmitted by $P_{\mathcal{Y}}$ can potentially reduce the chromatic number to the square root of its previous value, hence reduce the number of necessary bits to half its previous value (in a manner of speaking).

Lemma 8 For all (X, Y) pairs with $\hat{\mu}_{X|Y} \leq 2$ and all k, there is a 2-message protocol in which $P_{\mathcal{Y}}$ transmits k bits and $P_{\mathcal{X}}$ transmits at most $\lceil \frac{\hat{C}_1(X|Y)}{2^k} \rceil$ bits for all inputs.

Proof: According to the last lemma, there is a partition $S_1, \ldots, S_{2^k}$ of $S_{\mathcal{Y}}$ such that $\chi(G(X|Y)/_{S_i}) \leq \lceil \mathcal{G}^{\frac{1}{2^k}} \rceil$ for $i = 1, \ldots, 2^k$. $P_{\mathcal{X}}$ and $P_{\mathcal{Y}}$ agree on a k-bit encoding of each set S_i and on a $\lceil \mathcal{G}^{\frac{1}{2^k}} \rceil$-coloring of each $G(X|Y)/_{S_i}$.

$P_{\mathcal{Y}}$, given Y, uses k bits to transmit the code of the unique subset S_i that contains Y. Then $P_{\mathcal{X}}$ transmits $\lceil \log \lceil \mathcal{G}^{\frac{1}{2^k}} \rceil \rceil \leq \lceil \frac{\hat{C}_1(X|Y)}{2^k} \rceil$ bits describing the color of X in $G(X|Y)/_{S_i}$. □

The total number of bits transmitted in the lemma, $k + \lceil \frac{\hat{C}_1(X|Y)}{2^k} \rceil$, is a rapidly decreasing function of k. Therefore, to minimize transmissions, the recipient, $P_{\mathcal{Y}}$, transmits as many bits as possible: $\lceil \log \hat{C}_1(X|Y) \rceil$ bits and the informer, $P_{\mathcal{X}}$, sends only one bit, yielding Corollary 4.

5.2 $\hat{\mu}_{X|Y} > 2$: One Feedback Message is Almost Optimal

Lemma 7 showed that for any graph $G = (V, E)$ there is a partition E_1, E_2 of E such that both $\chi(G/_{E_1})$ and $\chi(G/_{E_2})$ are at most $\lceil\sqrt{\chi(G)}\rceil$. For hypergraphs, this is not generally true. Therefore the exact value of $\hat{C}_m(X|Y)$ for $m \geq 2$ may be difficult to determine, even if $\chi(G(X|Y))$ is known. Still we can show that $\hat{C}_2(X|Y)$, the two-message complexity of (X, Y), is never more than four times larger than the minimum complexity, $\hat{C}_\infty(X|Y)$.

We begin by showing that given any (X, Y) pair, we can construct a two-message protocol for (X, Y) that is close to optimal. The construction combines the characteristic-hypergraph coloring technique used in the short proof of Corollary 4 (given in Theorem 3) with a combinatorial result concerning perfect hash functions. Hash functions are discussed in Fredman and Komlós [9], Körner and Marton [10], and others.

Theorem 4 For all nontrivial (X, Y) pairs,

$$\hat{C}_2(X|Y) \leq \log\log\chi(G(X|Y)) + 3\log\hat{\mu}_{X|Y} + 4 \ .$$

Proof: Abbreviate $\chi \stackrel{\text{def}}{=} \mathcal{G}$. A collection $\mathcal{F}$ of functions, each defined over the set $\{1, \ldots, \chi\}$, *perfectly hashes* a collection $\mathcal{S}$ of subsets of $\{1, \ldots, \chi\}$, if for every subset $S \in \mathcal{S}$, there is a function $f \in \mathcal{F}$ that is one-to-one over S.

Let $C : S_\mathcal{X} \to \{1, \ldots, \chi\}$ be a χ-coloring of $G(X|Y)$. For every $y \in S_\mathcal{Y}$,

$$C(S_{X|Y}(y)) \stackrel{\text{def}}{=} \{C(x) : x \in S_{X|Y}(y)\}$$

is a subset of $\{1, \ldots, \chi\}$ containing all possible colors of X when $Y = y$. Define

$$\mathcal{S}_C \stackrel{\text{def}}{=} \{C(S_{X|Y}(y)) : y \in S_\mathcal{Y}\} \ ,$$

to be the collection of possible subsets of colors, and let b be an integer. We show that a collection $\mathcal{F}$ of functions from $\{1, \ldots, \chi\}$ to $\{1, \ldots, b\}$ that perfectly hashes $\mathcal{S}_C$ can be used to derive a two-message protocol for (X, Y) whose worst-case complexity is $\lceil\log|\mathcal{F}|\rceil + \lceil\log b\rceil$.

$P_\mathcal{X}$ and $P_\mathcal{Y}$ agree in advance on a $\lceil\log|\mathcal{F}|\rceil$-bit encoding of the functions in $\mathcal{F}$ and on a $\lceil\log b\rceil$-bit encoding of $\{1, \ldots, b\}$. When $P_\mathcal{X}$ is given X and $P_\mathcal{Y}$ is given Y, they execute the following protocol. $P_\mathcal{Y}$ finds a function $f_Y \in \mathcal{F}$ that perfectly hashes $C(S_{X|Y}(Y))$. Using $\lceil\log|\mathcal{F}|\rceil$ bits, he transmits the encoding of f_Y to $P_\mathcal{X}$. Now $P_\mathcal{X}$ knows X and f_Y. He computes $f_Y(C(X))$ and transmits it to $P_\mathcal{Y}$ using $\lceil\log b\rceil$ bits. Since f_Y is one-to-one over $C(S_{X|Y}(Y))$, $P_\mathcal{Y}$ can recover $C(X)$ from $f_Y(C(X))$. Since C is one-to-one over $S_{X|Y}(Y)$, $P_\mathcal{Y}$ can also recover X from $C(X)$.

To demonstrate a two-message protocol for (X, Y) with low worst-case complexity we therefore need to show the existence of a small collection $\mathcal{F}$ of functions with a small range $\{1,\ldots,b\}$ that perfectly hashes $\mathcal{S}_C$. Fredman and Komlós [9] showed that such *perfect-hash systems* exist. We simplify their proof and adapt it to our problem.

Pick a random function $F : \{1,\ldots,\chi\} \to \{1,\ldots,b\}$ by assigning, uniformly at random, a value in $\{1,\ldots,b\}$ to each element of $\{1,\ldots,\chi\}$. For all $y \in S_Y$, the probability that $C(S_{X|Y}(y))$ is hashed by F is:

$$\frac{b^{\underline{\mu}}}{b^{\mu}} \geq \frac{b^{\underline{\hat{\mu}}}}{b^{\hat{\mu}}}$$

where μ and $\hat{\mu}$ are abbreviations for $\mu_{X|Y}(y)$ and $\hat{\mu}_{X|Y}$, and $x^{\underline{i}} \stackrel{\mathrm{def}}{=} x \cdot (x-1) \cdot \ldots \cdot (x-i+1)$ denotes the ith falling power of x.

Independently pick m such random functions. The probability that some subset in $\mathcal{S}_C$ is not hashed by any of these functions is strictly smaller than:

$$\sum_{S \in \mathcal{S}_C} \left(1 - \frac{v(\mu)}{N(\mu)}\right)^m \leq |\mathcal{S}_C| \cdot \left(1 - \frac{b^{\underline{\hat{\mu}}}}{b^{\hat{\mu}}}\right)^m .$$

It follows that if

$$|\mathcal{S}_C| \cdot \left(1 - \frac{b^{\underline{\hat{\mu}}}}{b^{\hat{\mu}}}\right)^m \leq 1$$

then there is a family of m functions that hashes all subsets in $\mathcal{S}_C$. Therefore, for any

$$m \geq \frac{\log |\mathcal{S}_C|}{-\log(1 - \frac{b^{\underline{\hat{\mu}}}}{b^{\hat{\mu}}})}$$

there is a collection of m functions with range $\{1,\ldots,b\}$ that perfectly hashes $\mathcal{S}_C$. Choosing the smallest possible m, and replacing ceilings by an extra bit, the number of transmitted bits is at most:

$$\begin{aligned}
\log\left(\frac{\log |\mathcal{S}_C|}{-\log(1 - \frac{b^{\underline{\hat{\mu}}}}{b^{\hat{\mu}}})}\right) + \log b + 2 \quad &\leq \quad \log\left(\frac{b^{\hat{\mu}}}{b^{\underline{\hat{\mu}}}} \log |\mathcal{S}_C|\right) + \log b + 2 \\
&\leq \quad \log\log |\mathcal{S}_C| + \hat{\mu} \log \frac{b}{b - \hat{\mu}} + \log b + 2 .
\end{aligned}$$

Since the range $\{1,\ldots,b\}$ of the hash functions is given to our choice, we select $b = \hat{\mu}^2 + \hat{\mu}$ which minimizes the above expression, yielding:

$$\begin{aligned}
\hat{C}_2(X|Y) \quad &\leq \quad \log\log |\mathcal{S}_C| + \hat{\mu} \log \frac{\hat{\mu}^2 + \hat{\mu}}{\hat{\mu}^2} + \log(\hat{\mu}^2 + \hat{\mu}) + 2 \\
&\leq \quad \log\log |\mathcal{S}_C| + 2\log \hat{\mu} + \hat{\mu} \log \frac{\hat{\mu} + 1}{\hat{\mu}} + 3 \\
&\leq \quad \log\log |\mathcal{S}_C| + 2\log \hat{\mu} + 4 .
\end{aligned}$$

By definition, $\mathcal{S}_C$ is a collections of nonempty subsets of $\{1,\ldots,\chi\}$, each of size at most $\hat{\mu}$. Hence, $|\mathcal{S}_C| \le \chi^{\hat{\mu}}$, and the theorem follows. $\square$

In this paper we are mainly concerned with the possible discrepancy between the different $\hat{C}_m(X|Y)$'s and we combine the theorem with Corollary 3 and Theorem 2 to deduce that $\hat{C}_2(X|Y)$ is already very close to $\hat{C}_\infty(X|Y)$.

Corollary 5 For all (X,Y) pairs,

$$\hat{C}_2(X|Y) \le 4\hat{C}_\infty(X|Y) + 3\ . \qquad \square$$

6 Who Should Speak When I Want to Talk?

The optimal two message protocol for the league problem exhibits a peculiar phenomenon. To minimize total communication, the informant, $P_\mathcal{X}$, who wants to convey the winner, transmits just one bit while the recipient, $P_\mathcal{Y}$, who does not want to convey any information, transmits $\lceil \log\log t \rceil$ bits. Similarly, for the protocols upper bounding $\hat{C}_2(X|Y)$ in the last section, the informant transmits $\lceil \log b \rceil \le 2\log\hat{\mu}_{X|Y} + 2$ bits while the recipient transmits about $\log\log\mathcal{G} + \log\hat{\mu}_{X|Y}$ bits. If $\hat{\mu}_{X|Y} \ll \log\mathcal{G}$ the recipient transmits many more bits than the informant.

This of course does not imply that the recipient *must* transmit more bits than the informant. Every random pair has protocols in which the recipient does not transmit at all. In this section we show that for almost all random pairs with any given maximum ambiguity, none of these protocols is efficient. All optimal or near optimal protocols for these pairs require the recipient to transmit almost all the bits. The number of bits transmitted by the recipient cannot be appreciably reduced unless the informant transmits exponentially more bits than the minimum required. (These statements are gradually clarified throughout in this section.)

We restrict our discussion to (X,Y) pairs whose support set is contained in some finite product set $\mathcal{X}\times\mathcal{Y}$. Such a support set can be viewed as a collection of subsets $S_{X|Y}(y)$ of $\mathcal{X}$, one for each $y\in\mathcal{Y}$. If $S_{X|Y}(y) = S_{X|Y}(y')$ for some $y,y'\in\mathcal{Y}$ then all worst-case complexity measures of (X,Y) remain unchanged if we remove $\{(x,y) : x\in S_{X|Y}(y)\}$ from $S_{X,Y}$. Similarly, none of the worst-case complexity measures changes if we interchange $S_{X|Y}(y)$ and $S_{X|Y}(y')$ for some $y,y'\in\mathcal{Y}$. Therefore, from the point of view of worst-case complexities, (X,Y) pairs with $S_\mathcal{X}\subseteq\mathcal{X}$ fall into equivalence classes of pairs whose support sets correspond to the same collection of subsets of $\mathcal{X}$. There are $2^{|\mathcal{X}|}$ subsets of $\mathcal{X}$, hence $2^{2^{|\mathcal{X}|}}-1$ equivalence classes of "essentially identical[8] (X,Y) pairs" with $S_{\mathcal{X}} \subseteq \mathcal{X}$. It is in that sense of "random pairs" that we can talk of "most."

Of course, most random pairs have $\hat{\mu}_{X|Y} \geq \frac{|\mathcal{X}|}{2}$ and therefore $\hat{C}_\infty(X|Y) \geq \log|\mathcal{X}| - 1$. A straight-forward protocol that instructs $P_{\mathcal{X}}$ to transmit X is likely to be optimal for these pairs. This, however, does not hold for most of the "more interesting" (X,Y) pairs over $\mathcal{X} \times \mathcal{Y}$ having $\hat{\mu}_{X|Y} \ll |\mathcal{X}|$. More specifically, a random pair (X,Y) is *sparse* if

$$3\log\hat{\mu}_{X|Y} \ll \log\log|\mathcal{X}|$$

(roughly, $\hat{\mu}_{X|Y} \ll (\log|\mathcal{X}|)^{1/3}$). The league problem, for large t, is an extreme example of a sparse pair. Some sparse pairs, e.g., those with a cartesian product support set $S_{X,Y} \stackrel{\text{def}}{=} A \times B$ where $A \subseteq \mathcal{X}$ and $B \subseteq \mathcal{Y}$, have $\hat{C}_\infty(X|Y) = \hat{C}_1(X|Y)$, hence possess an optimal protocol in which $P_{\mathcal{X}}$ transmits all the bits. But, as we will soon see, these are a small minority.

Let ϕ be a protocol for a pair (X,Y). Define $\hat{L}_\phi^{\mathcal{X}}$ to be the supremum, over all inputs, of the number of bits transmitted by $P_{\mathcal{X}}$ under ϕ, and define $\hat{L}_\phi^{\mathcal{Y}}$ similarly. (A formal definition is given in the proof of Theorem 5 below.) We show that for most sparse (X,Y) pairs, $\hat{L}_\phi^{\mathcal{Y}}$, the number of bits transmitted by $P_{\mathcal{Y}}$ in the worst case, is about $\hat{C}_\infty(X|Y)$ in any optimal or near optimal protocol ϕ for (X,Y). $\hat{L}_\phi^{\mathcal{Y}}$ cannot be appreciably below $\hat{C}_\infty(X|Y)$ unless $\hat{L}_\phi^{\mathcal{X}}$, the number of bits transmitted by $P_{\mathcal{X}}$ in the worst case, is exponentially larger than $\hat{C}_\infty(X|Y)$.

The following "converse" of Lemma 8 underlies the results proven in this section. It establishes a non-symmetric tradeoff between the number of bits transmitted by $P_{\mathcal{X}}$ and $P_{\mathcal{Y}}$, showing that the bits transmitted by $P_{\mathcal{Y}}$ may be "more powerful."

Theorem 5 Let ϕ be a protocol for a random pair (X,Y). Then

$$\hat{L}_\phi^{\mathcal{X}} \geq \left\lceil \frac{\hat{C}_1(X|Y)}{2^{\hat{L}_\phi^{\mathcal{Y}}}} \right\rceil .$$

Proof: For $i \geq 0$, let $l_{\phi,i}^{\mathcal{X}}(x,y)$ denote the number of bits transmitted by $P_{\mathcal{X}}$ in messages $\sigma_1(x,y),\ldots,\sigma_i(x,y)$. For example, if $P_{\mathcal{X}}$ transmits the first message for (x,y), then $l_{\phi,i}^{\mathcal{X}}(x,y) = \sum_{j=1}^{\lceil i/2\rceil} |\sigma_{2j-1}(x,y)|$. Therefore, $l_{\phi,m_\phi(x,y)}^{\mathcal{X}}(x,y)$ is the number of bits transmitted by

[8] There can be other definitions of "essentially identical random pairs." Most likely, all yield results similar to those obtained in this section.

$P_{\mathcal{X}}$ for the input (x,y) and $\hat{L}_\phi^{\mathcal{X}} = \sup\{l^{\mathcal{X}}_{\phi,m_\phi(x,y)}(x,y) : (x,y) \in S_{X,Y}\}$ is the number of bit transmitted by $P_{\mathcal{X}}$ in the worst-case, as defined above. We define $l^{\mathcal{Y}}_{\phi,i}(x,y)$ similarly.

An induction similar to that of Theorem 2 can show that for all $i \geq 0$, there is an input (x,y) such that

$$\log \chi(G(\langle \sigma_j(x,y)\rangle_{j=1}^{i})) \geq \frac{\log \mathcal{G}}{2^{l^{\mathcal{Y}}_{\phi,i}(x,y)}} - l^{\mathcal{X}}_{\phi,i}(x,y) \ .$$

Without loss of generality, assume that the largest number of exchanged messages is finite. Letting i be that number of messages, we have $l^{\mathcal{X}}_{\phi,i}(x,y) = l^{\mathcal{X}}_{\phi,m_\phi(x,y)}(x,y)$, $l^{\mathcal{Y}}_{\phi,i}(x,y) = l^{\mathcal{Y}}_{\phi}(x,y)$, and $\chi(G(\langle \sigma_j(x,y)\rangle_{j=1}^{i})) = 1$. Hence

$$\hat{L}_\phi^{\mathcal{X}} \geq l^{\mathcal{X}}_{\phi,i}(x,y) \geq \frac{\log \mathcal{G}}{2^{l^{\mathcal{Y}}_{\phi,i}(x,y)}} \geq \frac{\log \mathcal{G}}{2^{\hat{L}_\phi^{\mathcal{Y}}}} \ .$$

The result then follows as $\hat{C}_1(X|Y) = \lceil \log \mathcal{G} \rceil$ according to Lemma 2. □

This shows that for every (X,Y) pair, and every protocol ϕ for it,

$$\hat{L}_\phi^{\mathcal{Y}} \geq \log \hat{C}_1(X|Y) - \log \hat{L}_\phi^{\mathcal{X}} \ .$$

Using Lemma 2, we can rewrite Theorem 4 as:

$$\hat{C}_\infty(X|Y) \leq \log \hat{C}_1(X|Y) + 3\log \mu + 4 \ .$$

Hence, for all (X,Y) pairs,

$$\hat{L}_\phi^{\mathcal{Y}} \geq \hat{C}_\infty(X|Y) - 3\log\mu - 4 - \log \hat{L}_\phi^{\mathcal{X}} \ . \tag{4}$$

The next lemma simplifies the analysis of this inequality. It shows that the complexities of most random pairs with any given maximum ambiguity fall within a narrow range. The lemma also states (at last) exactly what we mean by "most random pairs." The proof uses a well-known fact concerning the chromatic numbers of random graphs.

Lemma 9 Let $\mathcal{X}$ be a finite set and let $\mu \geq 2$ be an integer. A fraction of at least $1 - \frac{1}{|\mathcal{X}|^{\log\log|\mathcal{X}|}}$ of (X,Y) pairs with $S_{\mathcal{X}} \subseteq \mathcal{X}$ and $\hat{\mu}_{X|Y} = \mu$ have

$$\log|\mathcal{X}| - \log\log|\mathcal{X}| - 1 \leq \hat{C}_1(X|Y) \leq \log|\mathcal{X}|$$

and

$$\log\log|\mathcal{X}| \leq \hat{C}_\infty(X|Y) \leq \hat{C}_2(X|Y) \leq \log\log|\mathcal{X}| + 3\log\mu + 4 \ .$$

Proof: The upper bounds are given merely for completeness; we prove the lower bounds. A basic counting argument (Erdös and Spencer [11], or Palmer [12]) shows that a fraction of at most $\frac{1}{|\mathcal{X}|^{\log\log|\mathcal{X}|}}$ of the graphs on $|\mathcal{X}|$ vertices have an independent set larger than $2\log|\mathcal{X}|$ hence chromatic number smaller than $\frac{|\mathcal{X}|}{2\log|\mathcal{X}|}$. For $\mu > 2$, an even smaller fraction of hypergraphs with $\hat{\mu}_{X|Y} = \mu$ have chromatic number smaller than $\frac{|\mathcal{X}|}{2\log|\mathcal{X}|}$. Lemma 2 and Theorem 2 conclude the proof. $\square$

To interpret these results for sparse random pairs, assume that

$$3\log\mu \ll \log\log|\mathcal{X}|$$

(roughly, $\mu \ll (\log|\mathcal{X}|)^{1/3}$). The lemma implies that a fraction of at least $1 - \frac{1}{|\mathcal{X}|^{\log\log|\mathcal{X}|}}$ of (X, Y) pairs with $S_{\mathcal{X}} \subseteq \mathcal{X}$ and $\hat{\mu}_{X|Y} = \mu$ have

$$3\log\mu \ll \hat{C}_\infty(X|Y) \; .$$

For each (X, Y) pair in that majority, we can approximate Inequality (4) as

$$\hat{L}_\phi^{\mathcal{Y}} \gtrsim \hat{C}_\infty(X|Y) - \log \hat{L}_\phi^{\mathcal{X}} \; .$$

Therefore, for the above majority of sparse (X, Y) pairs, if ϕ is an optimal protocol for (X, Y), then $\hat{L}_\phi^{\mathcal{Y}}$, the number of bits transmitted by $P_{\mathcal{Y}}$ in the worst case, is $\gtrsim \hat{C}_\infty(X|Y) - \log\hat{C}_\infty(X|Y)$. Moreover, $\hat{L}_\phi^{\mathcal{Y}}$ can be appreciably lower than $\hat{C}_\infty(X|Y)$ only if ϕ is far worse than optimal: $\hat{L}_\phi^{\mathcal{X}}$, the number of bits transmitted by $P_{\mathcal{X}}$ in the worst case, is exponentially larger than $\hat{C}_\infty(X|Y)$.

7 Randomization and Errors: One-Way Communication

So far, we required that $P_{\mathcal{X}}$ and $P_{\mathcal{Y}}$ base their messages on preceding messages and the random variable they know. We also insisted that $P_{\mathcal{Y}}$ always know the correct value of X. In this section we relax both requirements. We allow the communicators to flip biased coins that influence both the choice of their messages and the value that $P_{\mathcal{Y}}$ assigns to X. We also let $P_{\mathcal{Y}}$ make ϵ error in deciding on the value of X.

Still, a variant of the protocol properties described in Section 2 must hold. The implicit-termination property is modified to require that at any given time the set of all possible messages be prefix free given the recipient's random variable and previous messages. The separate-transmissions and the correct-decision properties now say that the *bias* of the coin must be a function of previous messages and the random variable known to the transmitter of a message. Protocols complying with these requirements are *randomized* protocols.

We consider only one-way randomized protocols where $P_{\mathcal{X}}$ transmits a (random) message to $P_{\mathcal{Y}}$ based on which $P_{\mathcal{Y}}$ must (randomly) decide on the value of X. In so restricting ourselves we can avoid some of the subtleties in defining general randomized protocols.

Let Φ denote a randomized protocol. The number of bits transmitted when $P_{\mathcal{X}}$ knows x and $P_{\mathcal{Y}}$ knows y is now a random variable $L_\Phi(x,y)$; its expected value is $\bar{L}_\Phi(x,y)$. The *worst-case complexity* of Φ is:

$$\hat{L}_\Phi(X|Y) \stackrel{\text{def}}{=} \sup\{\bar{L}_\Phi(x,y) : (x,y) \in S_{X,Y}\} \ .$$

For every input $(x,y) \in S_{X,Y}$, the *assigned value*, $V_\Phi(x,y)$, which $P_{\mathcal{Y}}$ assumes is the value of X is a random variable and we let

$$\bar{E}_\Phi(x,y) \stackrel{\text{def}}{=} \Pr(V_\Phi(x,y) \neq x)$$

denote the probability that $P_{\mathcal{Y}}$ is in error for (x,y). The *error incurred by Φ in determining* X is:

$$\hat{E}_\Phi(X|Y) \stackrel{\text{def}}{=} \sup\{\bar{E}_\Phi(x,y) : (x,y) \in S_{X,Y}\} \ .$$

Note that the supremum in $\hat{L}_\Phi(X|Y)$ and $\hat{E}_\Phi(X|Y)$ is taken over the inputs. For each input, we average over the "coin flips." Finally, the *randomized one-way complexity* of (X,Y) with ϵ error is

$$\hat{C}_1^\epsilon(X|Y) \stackrel{\text{def}}{=} \inf\{\hat{L}_\Phi(X|Y) : \Phi \text{ is one-way and } \hat{E}_\Phi(X|Y) \leq \epsilon\}.$$

It is the minimum number of bits transmitted by a one-way randomized protocol ensuring that for all instances of (X,Y), $P_{\mathcal{Y}}$ determines X with probability of at least $1-\epsilon$.

Theorem 6 For all (X,Y) pairs and all $\epsilon > 0$,

$$\hat{C}_1^\epsilon(X|Y) \leq 2\left(\log\log\mathcal{G} + \log\hat{\mu}_{X|Y} + \log(1/\epsilon)\right) + 1 \ .$$

Proof: We use an hypergraph coloring argument to reduce the problem to one that can be solved with a variant of the prime-numbers algorithm derived by Rabin and Yao [13] to compute the *equality function*:

$$\text{eq}(i,j) = \begin{cases} 1 & \text{if } i = j; \\ 0 & \text{if } i \neq j. \end{cases}$$

As in El Gamal and Orlitsky [3], we modify the algorithm to convey a value rather than compute a function; but unlike the above citation, here, the value is a color.

Abbreviate $\chi \stackrel{\text{def}}{=} \mathcal{G}$ and $\mu \stackrel{\text{def}}{=} \hat{\mu}_{X|Y}$, and let $\sigma \stackrel{\text{def}}{=} \sqrt{2}\frac{\mu \ln \chi}{\epsilon}$. Then, $\chi > 60{,}000$ implies $\sigma > 31.1$ so the number of primes between σ and 2σ is at least $\frac{\sigma}{\sqrt{2}\ln\sigma}$ and at most $\frac{\sigma}{\ln\sigma}$. That is, $\frac{\sigma}{\sqrt{2}\ln\sigma} < \pi(2\sigma) - \pi(\sigma) < \frac{\sigma}{\ln\sigma}$. The number of primes between σ and 2σ that divide any positive integer $\leq \chi$ is always at most $\log_\sigma \chi$.

For an integer c, let $(c)_\alpha$ denote the value of $c \bmod \alpha$. $P_{\mathcal{X}}$ and $P_{\mathcal{Y}}$ agree on a χ-coloring $C : S_{\mathcal{X}} \to \{0, \ldots, \chi - 1\}$ of $G(X|Y)$. $P_{\mathcal{X}}$, given X, picks uniformly at random a prime α such that $\sigma < \alpha < 2\sigma$. He transmits α and $(C(X))_\alpha$ to $P_{\mathcal{Y}}$. $P_{\mathcal{Y}}$ constructs the set of colors

$$A(Y, \alpha, (C(X))_\alpha) \stackrel{\text{def}}{=} \{C(x) : x \in S_{X|Y}(Y) \text{ and } (C(x))_\alpha = (C(X))_\alpha\} \ .$$

The set is not empty as it contains $C(X)$. $P_{\mathcal{Y}}$ decides that $C(X)$ is a randomly chosen element of A. If $|A| = 1$, he decides on $C(X)$ correctly and, since C is a coloring of $G(X|Y)$, he also knows X. He can err only if $|A| \geq 2$. We show that this happens with probability smaller than ϵ for all inputs. For any $y \in S_{\mathcal{Y}}$ let $C_y \stackrel{\text{def}}{=} C(S_{X|Y}(y))$ be the set of possible colors of X when $Y = y$; clearly, $|C_y| = |S_{X|Y}(y)| \leq \mu$. Hence, for any input (x, y)

$$\begin{aligned}
\Pr(\text{ error } \mid (X,Y) = (x,y)) &< \Pr(|A(y, \alpha, (C(x))_\alpha)| \geq 2) \\
&= \Pr(\exists c' \in C_y : c' \neq C(x) \text{ and } (c')_\alpha = (C(x))_\alpha) \\
&\leq \sum_{c' \in C_y - \{C(x)\}} \Pr((c' - C(x))_\alpha = 0) \\
&= \sum_{c' \in C_y - \{C(x)\}} \frac{\text{number of primes in } \{\sigma, \ldots, 2\sigma\} \text{ dividing } (c' - c)}{\text{number of primes in } \{\sigma, \ldots, 2\sigma\}} \\
&\leq \sum_{c' \in C_y - \{C(x)\}} \frac{\log_\sigma \chi}{\sigma/(\sqrt{2}\ln\sigma)} \\
&\leq \sqrt{2}\mu\frac{\ln\chi}{\sigma} \\
&= \epsilon \ .
\end{aligned}$$

The number of bits required to transmit α is: $\log\frac{\sigma}{\ln\sigma} \leq \log\mu + \log\ln\chi + \log\frac{1}{\epsilon} + .5 - \log\ln 31.1$ bits. The expected number of bits required to transmit $C(X) \bmod \alpha$ is: $\overline{\log\alpha} \leq \log\overline{\alpha} \leq \log 1.7\sigma \leq \log\mu + \log\ln\chi + \log\frac{1}{\epsilon} + .5 + .77$ bits where $\overline{\alpha}$ is the expected value of α. Adding a bit to round up the above, we obtain:

$$\hat{\mathcal{C}}_1^\epsilon(X|Y) \leq 2\left(\log\ln\chi + \log\mu + \log\frac{1}{\epsilon}\right) + 1 \ . \qquad \square$$

We apply this theorem to relate $\hat{\mathcal{C}}_1^\epsilon(X|Y)$ to deterministic complexity measures.

Corollary 6 For all (X,Y) pairs and all $\epsilon > 0$,

$$\hat{\mathcal{C}}_1^\epsilon(X|Y) \leq 4\hat{C}_\infty(X|Y) + 2\log\frac{1}{\epsilon}\ . \qquad \square$$

Acknowledgements

I thank Toby Berger and Peter Doyle for a discussion concerning Theorem 2, and Don Coppersmith for helpful comments.

References

[1] H. Witsenhausen. The zero-error side information problem and chromatic numbers. *IEEE Transactions on Information Theory*, 22(5):592–593, September 1976.

[2] A. Orlitsky. Worst-case interactive communication II: Two messages are not optimal. *IEEE Transactions on Information Theory*, 37(4):995–1005, July 1991.

[3] A. El Gamal and A. Orlitsky. Interactive data compression. In *Proceedings of the 25th Annual Symposium on Foundations of Computer Science*, pages 100–108, 1984.

[4] C.H. Papadimitriou and M. Sipser. Communication complexity. In *Proceedings of the 14th Annual ACM Symposium on Theory of Computing*, pages 196–200, 1982.

[5] P. Duris, Z. Galil, and G. Schnitger. Lower bounds on communication complexity. In *Proceedings of the 16th Annual ACM Symposium on Theory of Computing*, pages 81–91, 1984.

[6] D. Slepian and J. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, 19:471–480, 1973.

[7] A. Orlitsky and A. El Gamal. Average and randomized communication complexity. *IEEE Transactions on Information Theory*, 36(1):3–16, January 1990.

[8] A. Orlitsky. *Communication Issues in Distributed Communication.* PhD thesis, Stanford University, 1986.

[9] M.L. Fredman and J. Komlós. On the size of separating systems and families of perfect hash functions. *SIAM Journal on Algebraic and Discrete Methods*, 5(1):61–68, 1984.

[10] J. Körner and K. Marton. New bounds for perfect hashing via information theory. *European Journal of Combinatorics*, 9:523–530, 1988.

[11] P. Erdös and J. Spencer. *Probabilistic Methods in Combinatorics.* Academic Press, 1974.

[12] E.M. Palmer. *Graphical Evolution: An Introduction to the Theory of Random Graphs.* John Wiley & Sons, 1985.

[13] M.O. Rabin and A.C. Yao. Unpublished manuscript, 1979. See Yao [**?**].