The Entropic Uncertainty Principle and the Fast Fourier Transform

Charles S. Peskin
Courant Institute of Mathematical Sciences, New York University
Modeling and Simulation Group Seminar — March 5, 2020

Abstract

The entropic uncertainty principle for the discrete Fourier transform states that $H(u) + H(F_n u) \geq \log(n)$, where F_n is the discrete Fourier transform of order n, and $H(u)$ is the entropy of the discrete probability distribution given by $P_j = (|u_j|/\|u\|)^2$. This is a special case of a known result [1] that requires substantial mathematical apparatus for its proof. Here, we give an elementary proof for the special case of the discrete Fourier transform with $n = 2^p$. Our method of proof is induction on p. By proceeding in this way, we reveal a connection between the entropic uncertainty principle and the Fast Fourier Transform algorithm.

Introduction

The discrete Fourier transform $F_n \colon \mathbb{C}^n \to \mathbb{C}^n$ is defined by

$$(F_n u)_k = \frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} \mathrm{e}^{-i\frac{2\pi}{n}jk} u_j, \tag{1}$$

for $k = 0, 1, \ldots, n-1$. It is easy to check that F_n is unitary:

$$\|F_n u\| = \|u\|, \tag{2}$$

where $\|\ \ \|$ is the Euclidean norm.

The entropy of a vector $u \in \mathbb{C}^n$ may be defined as follows:

$$H(u) = -\sum_{j=0}^{n-1} \left(\frac{|u_j|}{\|u\|}\right)^2 \log\left(\left(\frac{|u_j|}{\|u\|}\right)^2\right). \tag{3}$$

Note that $H(u)$ is invariant under any permutation of the components of u, and also that H is homogeneous of degree 0, that is,

$$H(cu) = H(u) \tag{4}$$

for any complex number c. If we interpret $P_j = (|u_j|/\|u\|)^2$ as the probability of the event j, for $j = 0, 1, \ldots, n-1$, then $H(u)$ is the entropy of the discrete probability distribution $(P_0, \ldots, P_{n-1})$.

It is a known theorem [1] that

$$H(u) + H(F_n u) \geq \log(n). \tag{5}$$

This is called the *entropic uncertainty principle* for the discrete Fourier transform.[1] The proof of (5) in [1] is very short and elegant, and in fact it considers a more general case, as explained below, but it makes use of concepts that seem more advanced than necessary for the problem at hand, which is concerned only with finite-dimensional vector spaces. That is one motivation for providing a new proof here. Another is that the lower bound $\log(n)$ of the entropic uncertainty principle is strongly suggestive of the operation count $O(n\log(n))$ of the Fast Fourier Transform (FFT) algorithm, and this suggests a method of proof at least for the special case $n = 2^p$ (which is the only case that we shall consider here) in which each of p steps contributes $\log(2)$ to the total entropy. That is indeed how we shall proceed.

As mentioned above, the theorem proved in [1] is actually more general than (5). It may be stated (in the finite-dimensional case) as follows. Let U be any unitary $n \times n$ matrix, let $M = \max_{jk} |U_{jk}|$, and let x be any vector in $\mathbb{C}^n$. Then

$$H(x) + H(Ux) \geq 2\log\left(\frac{1}{M}\right). \tag{6}$$

Since each column (or row) of a unitary matrix is a unit vector, the constant M in this inequality must lie in the interval $[\frac{1}{\sqrt{n}}, 1]$. We get the strongest result, then,

[1] We have not been specific about the base of the logarithm because the entropic uncertainty principle is valid in any base, provided of course that the same base is used in the definition of the entropy as in the statement of the uncertainty principle. In the proof that follows (in particular in the Appendix, where derivatives of the logarithm are involved), it is most convenient to use base e, and that is how we shall proceed. Note, however, that when $n = 2^p$, the result takes its simplest form if the logarithm used is base 2, since in that case the lower bound of the total entropy is simply p.

when M is at the lower end of this interval, so that the constant on the right-hand side of (6) is as large as possible, and in fact is equal to $\log(n)$. This is actually the case for the discrete Fourier transform, which therefore is not just any example, but rather the most extreme example possible, in the sense that its entropic uncertainty principle is as strong as it could possibly be. This is perhaps a further motivation for the separate consideration of the discrete Fourier transform.

An entropy inequality for each step of the FFT

We start by deriving the recusion relation that is fundamental to the Fast Fourier Transform algorithm in the special case $n = 2^p$. In that case, n is even for $p > 0$, and we have

$$\begin{aligned}(F_n u)_k &= \frac{1}{\sqrt{n}} \sum_{j=0}^{\frac{n}{2}-1} \mathrm{e}^{-i\frac{2\pi}{n}(2j)k} u_{2j} + \frac{1}{\sqrt{n}} \sum_{j=0}^{\frac{n}{2}-1} \mathrm{e}^{-i\frac{2\pi}{n}(2j+1)k} u_{2j+1} \\ &= \frac{1}{\sqrt{2}} \frac{1}{\sqrt{n/2}} \sum_{j=0}^{\frac{n}{2}-1} \mathrm{e}^{-i\frac{2\pi}{n/2}jk} u_{2j} + \frac{1}{\sqrt{2}} \frac{\mathrm{e}^{-i\frac{2\pi}{n}k}}{\sqrt{n/2}} \sum_{j=0}^{\frac{n}{2}-1} \mathrm{e}^{-i\frac{2\pi}{n/2}jk} u_{2j+1}. \end{aligned} \tag{7}$$

Note that $(F_n u)_{k+n/2}$ is exactly the same as $(F_n u)_k$, except that

$$\mathrm{e}^{-i\frac{2\pi}{n}\left(k+\frac{n}{2}\right)} = \mathrm{e}^{-i\frac{2\pi}{n}k}\mathrm{e}^{-i\pi} = -\mathrm{e}^{-i\frac{2\pi}{n}k}. \tag{8}$$

Thus, if we restrict k to $\{0, 1, \ldots, \frac{n}{2} - 1\}$ and define

$$v_k^0 = (F_n u)_k, \tag{9}$$
$$v_k^1 = (F_n u)_{k+n/2}, \tag{10}$$

then we may write (7) more succinctly as

$$v^0 = \frac{1}{\sqrt{2}} \left(F_{n/2} u^{\text{even}} + D F_{n/2} u^{\text{odd}} \right), \tag{11}$$
$$v^1 = \frac{1}{\sqrt{2}} \left(F_{n/2} u^{\text{even}} - D F_{n/2} u^{\text{odd}} \right). \tag{12}$$

In this pair of equations,

$$u_j^{\text{even}} = u_{2j}, \tag{13}$$
$$u_j^{\text{odd}} = u_{2j+1}, \tag{14}$$

for $j = 0, 1, \ldots, \frac{n}{2} - 1$, and D is an $\frac{n}{2} \times \frac{n}{2}$ matrix with off-diagonal elements equal to zero and with diagonal elements given by

$$D_{kk} = \mathrm{e}^{-i\frac{2\pi}{n}k}, \tag{15}$$

for $k = 0, 1, \ldots, \frac{n}{2} - 1$. Note that D is unitary.

Let

$$w^0 = F_{n/2} u^{\text{even}}, \tag{16}$$

$$w^1 = F_{n/2} u^{\text{odd}}. \tag{17}$$

Then (11-12) become

$$v^0 = \frac{1}{\sqrt{2}} \left(w^0 + D w^1 \right), \tag{18}$$

$$v^1 = \frac{1}{\sqrt{2}} \left(w^0 - D w^1 \right), \tag{19}$$

or, in components

$$v^0_k = \frac{1}{\sqrt{2}} \left(w^0_k + D_{kk} w^1_k \right), \tag{20}$$

$$v^1_k = \frac{1}{\sqrt{2}} \left(w^0_k - D_{kk} w^1_k \right), \tag{21}$$

for $k = 0, 1, \ldots, \frac{n}{2} - 1$. Note that there is no coupling in equations (20-21) between different values of k.

Since the entropy of a vector as defined above only involves the absolute values of the components, we evaluate

$$|v^0_k|^2 = \frac{1}{2} \left(|w^0_k|^2 + |w^1_k|^2 + q_k \right), \tag{22}$$

$$|v^1_k|^2 = \frac{1}{2} \left(|w^0_k|^2 + |w^1_k|^2 - q_k \right), \tag{23}$$

where

$$q_k = w^0_k \overline{D_{kk} w^1_k} + \overline{w^0_k} D_{kk} w^1_k. \tag{24}$$

Note that q_k is real, and also that

$$|v^0_k|^2 + |v^1_k|^2 = |w^0_k|^2 + |w^1_k|^2. \tag{25}$$

Let v and w be vectors in $\mathbb{C}^n$ defined by

$$v_k = v_k^0, \qquad v_{k+n/2} = v_k^1, \tag{26}$$

$$w_k = w_k^0, \qquad w_{k+n/2} = w_k^1, \tag{27}$$

for $k = 0, 1, \ldots, \frac{n}{2} - 1$. Note that

$$\|v\| = \|w\|. \tag{28}$$

The goal of this section is to derive a lower bound on $H(v) + H(w)$. We have

$$H(w) = -\sum_{k=0}^{\frac{n}{2}-1} \left[\left(\frac{|w_k^0|}{\|w\|} \right)^2 \log \left(\left(\frac{|w_k^0|}{\|w\|} \right)^2 \right) + \left(\frac{|w_k^1|}{\|w\|} \right)^2 \log \left(\left(\frac{|w_k^1|}{\|w\|} \right)^2 \right) \right], \tag{29}$$

and it is straightforward to check (see Appendix 2) that this expression for $H(w)$ can be rewritten in the following equivalent way

$$H(w) = H_0(w) + \sum_{k=0}^{\frac{n}{2}-1} \frac{|w_k^0|^2 + |w_k^1|^2}{\|w\|^2} H\left(\left(w_k^0, w_k^1 \right) \right), \tag{30}$$

where

$$H_0(w) = -\sum_{k=0}^{\frac{n}{2}-1} \frac{|w_k^0|^2 + |w_k^1|^2}{\|w\|^2} \log \left(\frac{|w_k^0|^2 + |w_k^1|^2}{\|w\|^2} \right), \tag{31}$$

and where $H\left((w_k^0, w_k^1) \right)$ denotes the entropy of the 2-vector with components w_k^0 and w_k^1, which is

$$\begin{aligned} H\left(\left(w_k^0, w_k^1 \right) \right) = &- \frac{|w_k^0|^2}{|w_k^0|^2 + |w_k^1|^2} \log \left(\frac{|w_k^0|^2}{|w_k^0|^2 + |w_k^1|^2} \right) \\ &- \frac{|w_k^1|^2}{|w_k^0|^2 + |w_k^1|^2} \log \left(\frac{|w_k^1|^2}{|w_k^0|^2 + |w_k^1|^2} \right). \end{aligned} \tag{32}$$

Similarly,

$$H(v) = H_0(v) + \sum_{k=0}^{\frac{n}{2}-1} \frac{|v_k^0|^2 + |v_k^1|^2}{\|v\|^2} H\left(\left(v_k^0, v_k^1 \right) \right). \tag{33}$$

Because of (25-28), we can rewrite much of the expression for $H(v)$ in terms of w. In fact,

$$H(v) = H_0(w) + \sum_{k=0}^{\frac{n}{2}-1} \frac{|w_k^0|^2 + |w_k^1|^2}{\|w\|^2} H\left(\left(v_k^0, v_k^1 \right) \right). \tag{34}$$

Now adding equations (30) & (34), we get

$$H(v) + H(w) = 2H_0(w) + \sum_{k=0}^{\frac{n}{2}-1} \frac{|w_k^0|^2 + |w_k^1|^2}{\|w\|^2} \left[H\left(\left(v_k^0, v_k^1\right)\right) + H\left(\left(w_k^0, w_k^1\right)\right)\right]. \tag{35}$$

We shall prove in Appendix 1 that

$$H\left(\left(v_k^0, v_k^1\right)\right) + H\left(\left(w_k^0, w_k^1\right)\right) \geq \log(2). \tag{36}$$

It then follows from this and (35) that

$$H(v) + H(w) \geq \log(2) + 2H_0(w). \tag{37}$$

We shall see in the next section how this lower bound can be used in an inductive proof of the entropic uncertainty principle.

Induction proof of the entropic uncertainty principle

We are now ready to prove (5) in the special case that n is a power of 2. First, we note that (5) is trivially true for $n = 1$, since $H(u)$ and $H(F_1 u)$ are both equal to zero in that case. Next, we make the induction hypothesis that

$$H(x) + H(F_{n/2}x) \geq \log\left(\frac{n}{2}\right), \tag{38}$$

for all $x \in \mathbb{C}^{n/2}$. Setting $x = u^{\text{even}}$ and then $x = u^{\text{odd}}$, we get

$$H(u^{\text{even}}) + H(w^0) \geq \log\left(\frac{n}{2}\right), \tag{39}$$

$$H(u^{\text{odd}}) + H(w^1) \geq \log\left(\frac{n}{2}\right), \tag{40}$$

since $w^0 = F_{n/2}u^{\text{even}}$ and $w^1 = F_{n/2}u^{\text{odd}}$.

Now recall (see equations (13-14) and (27)) that the components of $u \in \mathbb{C}^n$ are those of u^{even} and those of u^{odd}, with $\|u\|^2 = \|u^{\text{even}}\|^2 + \|u^{\text{odd}}\|^2$, and similarly, that the components of $w \in \mathbb{C}^n$ are those of w^0 and those of w^1, with $\|w\|^2 = \|w^0\|^2 + \|w^1\|^2$. It is then straightforward to check that

$$H(u) = \frac{\|u^{\text{even}}\|^2}{\|u\|^2} H(u^{\text{even}}) + \frac{\|u^{\text{odd}}\|^2}{\|u\|^2} H(u^{\text{odd}}) + H\left(\left(\|u^{\text{even}}\|, \|u^{\text{odd}}\|\right)\right), \tag{41}$$

$$H(w) = \frac{\|w^0\|^2}{\|w\|^2} H(w^0) + \frac{\|w^1\|^2}{\|w\|^2} H(w^1) + H\left(\left(\|w^0\|, \|w^1\|\right)\right). \tag{42}$$

Now recall that $F_{n/2}$ is unitary, and that $w^0 = F_{n/2}u^{\text{even}}$ and $w^1 = F_{n/2}u^{\text{odd}}$. Thus, $\|u^{\text{even}}\| = \|w^0\|$, $\|u^{\text{odd}}\| = \|w^1\|$, and $\|u\| = \|w\|$. Therefore, if we add equations (41) & (42), the result can be written as

$$\begin{aligned} H(u) + H(w) =& \frac{\|w^0\|^2}{\|w\|^2}\left(H(u^{\text{even}}) + H(w^0)\right) \\ &+ \frac{\|w^1\|^2}{\|w\|^2}\left(H(u^{\text{odd}}) + H(w^1)\right) + 2H\left(\left(\|w^0\|, \|w^1\|\right)\right). \end{aligned} \tag{43}$$

Applying the inequalities (39-40) to the above equation gives

$$H(u) + H(w) \geq \log\left(\frac{n}{2}\right) + 2H\left(\left(\|w^0\|, \|w^1\|\right)\right). \tag{44}$$

Now if we add the inequalities (37) & (44), we get

$$H(u) + H(v) + 2H(w) \geq \log\left(\frac{n}{2}\right) + \log(2) + 2H_0(w) + 2H\left(\left(\|w^0\|, \|w^1\|\right)\right). \tag{45}$$

Of course, $\log\left(\frac{n}{2}\right) + \log(2) = \log(n)$, which is the lower bound on the entropic entropy that we are trying to prove. Also, with the help of (42) we can rewrite (45) in the following way:

$$H(u) + H(v) \geq \log(n) + 2Q(w), \tag{46}$$

where

$$Q(w) = H_0(w) - \frac{\|w^0\|^2}{\|w\|^2}H(w^0) - \frac{\|w^1\|^2}{\|w\|^2}H(w^1) \tag{47}$$

To complete the proof, we need only show that $Q(w) \geq 0$. Let

$$P_k^0 = \frac{|w_k^0|^2}{\|w^0\|^2}, \quad P_k^1 = \frac{|w_k^1|^2}{\|w^1\|^2}, \tag{48}$$

$$P_k = \frac{\|w^0\|^2}{\|w\|^2}P_k^0 + \frac{\|w^1\|^2}{\|w\|^2}P_k^1 = \frac{|w_k^0|^2 + |w_k^1|^2}{\|w\|^2}. \tag{49}$$

Further, let

$$\alpha_0 = \frac{\|w^0\|^2}{\|w\|^2} \geq 0, \quad \alpha_1 = \frac{\|w^1\|^2}{\|w\|^2} \geq 0, \tag{50}$$

and note that $\alpha_0 + \alpha_1 = 1$.

It is then easy to check that

$$H(w^0) = \mathcal{H}(P^0), \tag{51}$$
$$H(w^1) = \mathcal{H}(P^1), \tag{52}$$
$$H_0(w) = \mathcal{H}(\alpha_0 P^0 + \alpha_1 P^1), \tag{53}$$

where $\mathcal{H}(P)$ is the entropy of the discrete probability distribution P. Therefore,

$$Q(w) = \mathcal{H}(\alpha_0 P^0 + \alpha_1 P^1) - \alpha_0 \mathcal{H}(P^0) - \alpha_1 \mathcal{H}(P^1), \tag{54}$$

which is clearly non-negative because of the well-known convexity of $-\mathcal{H}$. Thus, (46) implies that

$$H(u) + H(v) \geq \log(n). \tag{55}$$

Since $v = F_n u$, this completes the induction step, and hence the proof.

References

[1] Dembo A, Cover TM, and Thomas JA 1991: Information Theoretic Inequalities. **IEEE Transactions on Information Theory 37**(6):1501-1517, see Theorem 23 on page 1513.

Appendix 1

Let v_0, v_1, w_0, w_1 be complex numbers related by

$$v_0 = \frac{1}{\sqrt{2}} \left(w_0 + D w_1 \right), \tag{56}$$
$$v_1 = \frac{1}{\sqrt{2}} \left(w_0 - D w_1 \right), \tag{57}$$

where D is a complex number with $|D| = 1$. Let $v = (v_0, v_1)$, $w = (w_0, w_1)$. We claim that

$$H(v) + H(w) \geq \log(2). \tag{58}$$

Here,

$$\begin{aligned} H(w) = &- \frac{|w_0|^2}{|w_0|^2 + |w_1|^2} \log \left(\frac{|w_0|^2}{|w_0|^2 + |w_1|^2} \right) \\ &- \frac{|w_1|^2}{|w_0|^2 + |w_1|^2} \log \left(\frac{|w_1|^2}{|w_0|^2 + |w_1|^2} \right), \end{aligned} \tag{59}$$

and similarly for $H(v)$.

In the following, we shall think of $H(v) + H(w)$ as a function of w, since v is determined by w through (56-57). From (56-57), we have

$$|v_0|^2 = \frac{1}{2}\left(|w_0|^2 + |w_1|^2 + q\right), \tag{60}$$

$$|v_1|^2 = \frac{1}{2}\left(|w_0|^2 + |w^1|^2 - q\right), \tag{61}$$

where

$$q = w_0\overline{Dw_1} + \overline{w_0}Dw_1. \tag{62}$$

Note that q is real, and also that

$$|v_0|^2 + |v_1|^2 = |w_0|^2 + |w_1|^2. \tag{63}$$

Because of the homogeneity of H we can, without loss of generality, assume that

$$|w_0|^2 + |w_1|^2 = 1. \tag{64}$$

It follows, of course, that $|v_0|^2 + |v_1|^2 = 1$, so

$$\begin{aligned} H(v) + H(w) = &- |v_0|^2 \log\left(|v_0|^2\right) - |v_1|^2 \log\left(|v_1|^2\right) \\ &- |w_0|^2 \log\left(|w_0|^2\right) - |w_1|^2 \log\left(|w_1|^2\right). \end{aligned} \tag{65}$$

Also, (60-61) become

$$|v_0|^2 = \frac{1}{2}(1+q), \tag{66}$$

$$|v_1|^2 = \frac{1}{2}(1-q), \tag{67}$$

so

$$H(v) + H(w) = f(q) - |w_0|^2 \log\left(|w_0|^2\right) - |w_1|^2 \log\left(|w_1|^2\right), \tag{68}$$

where

$$f(q) = -\left(\frac{1+q}{2}\right)\log\left(\frac{1+q}{2}\right) - \left(\frac{1-q}{2}\right)\log\left(\frac{1-q}{2}\right). \tag{69}$$

The function $f(q)$ plays an important role in the following, so we note some of its properties. First $f(q)$ is defined on the interval $[-1, 1]$. Its values at the endpoints of this interval, which are defined by taking limits from inside the interval, are

$$f(\pm 1) = 0. \tag{70}$$

The function f is even:

$$f(q) = f(-q), \tag{71}$$

and its first and second derivatives are given by

$$f'(q) = \frac{1}{2}\log\left(\frac{1-q}{1+q}\right), \tag{72}$$

$$f''(q) = -\frac{1}{1-q^2} < 0. \tag{73}$$

Since $f''(q) < 0$, f has a unique maximum which occurs where $f'(q) = 0$, namely at $q = 0$. The maximum value of f is given by

$$f(0) = \log(2). \tag{74}$$

Also, since $f'(q) < 0$ for $q > 0$, and since f is an even function, we may conclude that f is a decreasing function of $|q|$.

Our task now is to minimize $H(v) + H(w)$, and to show that the minimum value of $H(v) + H(w)$ is $\log(2)$. [2] The minimization will be done by a two-step procedure. In the first step, we fix $|w_0|$ and $|w_1|$, and minimize with respect to q. Then, with q set equal to its optimal value as a function of $|w_0|$ and $|w_1|$, we minimize with respect to $|w_0|$ and $|w_1|$, subject, of course, to the constraint that $|w_0|^2 + |w_1|^2 = 1$.

With $|w_0|$ and $|w_1|$ fixed, $H(w)$ is fixed, so our task is simply to minimize $H(v) = f(q)$. This is done by making $|q|$ as large as possible. From (62), we have

$$|q| \leq 2|w_0||w_1|, \tag{75}$$

with equality if

$$D\frac{w_1}{|w_1|} = \pm\frac{w_0}{|w_0|}, \tag{76}$$

which can certainly be achieved with $|w_0|$ and $|w_1|$ fixed merely by adjusting the angles of w_0 and w_1 in the complex plane (recall that $|D| = 1$). Thus the optimal value of q is given by

$$|q| = \pm 2|w_0||w_1|, \tag{77}$$

and the sign does not matter, since f is an even function. With this optimal choice of q, our task is reduced to minimizing

$$H(v) + H(w) = f(2|w_0||w_1|) - |w_0|^2\log\left(|w_0|^2\right) - |w_1|^2\log\left(|w_1|^2\right), \tag{78}$$

[2] Note that this is the *maximum* value of f, a fact that does not seem helpful at this point, since we are seeking the *minimum* value of $H(v) + H(w) = f(q) + H(w)$.

subject to the constraint $|w_0|^2 + |w_1|^2 = 1$. We can build in this constraint by setting

$$|w_0| = \cos\theta, \tag{79}$$
$$|w_1| = \sin\theta, \tag{80}$$

so that

$$2|w_0||w_1| = \sin(2\theta), \tag{81}$$
$$|w_0|^2 = \cos^2\theta = \frac{1}{2}(1 + \cos(2\theta)), \tag{82}$$
$$|w_1|^2 = \sin^2\theta = \frac{1}{2}(1 - \cos(2\theta)). \tag{83}$$

With these substitutions, we find the beautiful result that

$$H(v) + H(w) = f(\sin(2\theta)) + f(\cos(2\theta)). \tag{84}$$

Let

$$g(\theta) = f(\sin(2\theta)) + f(\cos(2\theta)). \tag{85}$$

Our task is to minimize $g(\theta)$. Since f is even, g is even. Also since f is even, it is easily checked that $g(\theta + \pi/4) = g(\theta)$. Thus, we may restrict consideration to the interval $[0, \pi/8]$. We claim that g is an increasing function on this interval. This may be shown by noting that

$$g'(\theta) = \sin(2\theta)\cos(2\theta)\left(A(\cos(2\theta)) - A(\sin(2\theta))\right), \tag{86}$$

where

$$A(x) = \frac{1}{x}\log\left(\frac{1+x}{1-x}\right) \tag{87}$$

is an increasing function for $x \in (0, 1)$, see below. On the interval $[0, \pi/8]$, we have

$$0 \leq \sin(2\theta) \leq \frac{1}{\sqrt{2}} \leq \cos(2\theta), \tag{88}$$

and therefore, since A is increasing, $g'(\theta) \geq 0$. Thus, the minimum of g occurs at $\theta = 0$, where we have

$$g(0) = f(0) + f(1) = \log(2) + 0 = \log(2). \tag{89}$$

It follows, as claimed, that $H(v) + H(w) \geq \log(2)$.

For completeness, we give a proof that $A(x)$ is increasing for $x \in (0,1)$. We have

$$A'(x) = \frac{1}{x^2}\left(K(x) - L(x)\right), \tag{90}$$

where

$$K(x) = \frac{2x}{1-x^2}, \tag{91}$$

$$L(x) = \log\left(\frac{1+x}{1-x}\right). \tag{92}$$

Note that $K(0) = L(0) = 0$. Also,

$$K'(x) = \frac{2\left(1+x^2\right)}{\left(1-x^2\right)^2}, \tag{93}$$

$$L'(x) = \frac{2}{1-x^2} = \frac{2\left(1-x^2\right)}{\left(1-x^2\right)^2}, \tag{94}$$

so $K'(x) > L'(x)$ for $x \in (0,1)$. It follows, since $K(0) = L(0)$, that $K(x) > L(x)$ for $x \in (0,1)$, and hence that $A'(x) > 0$ on this interval.

Appendix 2

In this appendix we consider the behavior of entropy under the aggregation of outcomes into disjoint sets of outcomes. Several of the algebraic manipulations of entropy in the main text of this paper are special cases of the following.

Let p_i be a discrete probability distribution, and let the possible outcomes i be partitioned into disjoint sets C_m, so that every i belongs to exactly one of the sets C_m. Let

$$P_m = \sum_{i \in C_m} p_i, \tag{95}$$

so that P_m is the probability of an outcome in C_m. Then

$$\begin{aligned}
\mathcal{H}(p) &= -\sum_i p_i \log p_i, \\
&= -\sum_m \sum_{i \in C_m} p_i \log p_i, \\
&= -\sum_m P_m \sum_{i \in C_m} \frac{p_i}{P_m} \left(\log \frac{p_i}{P_m} + \log P_m \right), \\
&= -\sum_m P_m \sum_{i \in C_m} \frac{p_i}{P_m} \log \frac{p_i}{P_m} - \sum_m P_m \log P_m, \\
&= \sum_m P_m \mathcal{H}_m \quad + \quad \mathcal{H}_0.
\end{aligned} \tag{96}$$

In the last line of the above equation, $\mathcal{H}_m = -\sum_{i \in C_m} \frac{p_i}{P_m} \log \frac{p_i}{P_m}$ is the conditional entropy of the collection of outcomes C_m, given that the outcome is in that collection, and $\mathcal{H}_0 = \mathcal{H}(P) = -\sum_m P_m \log P_m$ is the entropy of the aggregated probability distribution P.

The above result is used in two different ways in this paper. In one version we aggregate n outcomes into $n/2$ collections each of which has two members, and in the other case into two collections of outcomes each of which has $n/2$ members.