Society for the Study of Economic Inequality

Working Paper Series

Shannon-Theil-Rawls: Information Theory, Inequality and the Veil of Ignorance

Ravi Kanbur

ECINEQ 2024 669

ECINEQ 2024 669
2024 March
www.ecineq.org

Shannon-Theil-Rawls: Information Theory, Inequality and the Veil of Ignorance

Ravi Kanbur
Cornell University

Abstract

This paper shows the power of applying Shannon's (1948) information theory perspective to inequality measurement by considering the thought experiment of drawing a dollar at random from an income distribution and asking who the dollar came from. The surprise at being told who the dollar came from, and the task of designing a set of questions with yes/no answers which will get us to the person, are two sides of the same coin but with interesting interpretations. The Theil index of inequality, which Theil (1967) himself derived with reference to information theory and entropy but did not then explore further, is shown to have interpretations beyond its simple Daltonian properties such as satisfying the principle of transfers or being sub-group decomposable. It can be interpreted as a statistical test of the hypothesis of fairness, and as a quantitative measure of the difficulty of achieving Rawls's (1971) original position behind the veil of ignorance.

Keyword: Information Theory, Inequality, Veil of Ignorance

JEL Cassification: D31, D63

Shannon-Theil-Rawls:
Information Theory, Inequality and the Veil of Ignorance*

By

Ravi Kanbur
www.kanbur.dyson.cornell.edu

This version: 20 January 2024

Contents

1. Introduction
2. Daltonian Inequality Measurement
3. Who Did That Dollar Come From?
4. Priors, Posteriors and Inequality
5. Questions and Codes
6. Codes, Labels and the Veil of Ignorance
7. Conclusion

   References

Abstract

This paper shows the power of applying Shannon's (1948) information theory perspective to inequality measurement by considering the thought experiment of drawing a dollar at random from an income distribution and asking who the dollar came from. The surprise at being told who the dollar came from, and the task of designing a set of questions with yes/no answers which will get us to the person, are two sides of the same coin but with interesting interpretations. The Theil index of inequality, which Theil (1967) himself derived with reference to information theory and entropy but did not then explore further, is shown to have interpretations beyond its simple Daltonian properties such as satisfying the principle of transfers or being sub-group decomposable. It can be interpreted as a statistical test of the hypothesis of fairness, and as a quantitative measure of the difficulty of achieving Rawls's (1971) original position behind the veil of ignorance.

* Paper based on keynote presentation at LACEA-LAMES conference, Puebla, 2019, and presentation at conference in honor of Joseph Stiglitz's 80th birthday, Milan, 2023. I am grateful to conference participants, and to Duncan Foley, Sanjay Reddy and Paulo dos Santos, for helpful discussions.

1. Introduction

Information Theory a la Shannon was introduced to economics by Theil (1967) in his book. *Economics and Information Theory*. Chapter 4 of the book was entitled "The Measurement of Inequality" and introduced the now famous Theil Entropy Measure of Inequality. The next quarter century saw a flurry of activity in this area, for example in the work of Cowell (1980) and Shorrocks (1980), developing the properties of what became known as the Generalized Entropy Family of Inequality Measures. This work focused in particular on the sub-group decomposability of these measures, a property which proves to be operationally very useful.

However, even though the label "generalized entropy measures of inequality" is prevalent in empirical work on inequality, and despite notable exceptions[1], the direct engagement with entropy and information theory in discussions of inequality seems to have fallen by the wayside in mainstream economics. I would like to use this platform of a conference in honor of Joseph Stiglitz's 80th birthday to revisit, to revive, and I hope to reinvigorate the link between information theory and inequality. In doing so I will also attempt to relate quantitative inequality measurement to the discourse on the Rawlsian Veil of Ignorance.

The conventional approach to normative inequality measurement is encapsulated in Dalton's (1920) specification of normative criteria that measures of inequality should satisfy, such as the principle of transfers. This historical development is traced in Section 2 of the paper. In section 3, I argue that information theory presents a distinctive normative perspective on inequality, when presented as seeking the source of a dollar drawn at random from an income distribution—the more difficult (easier) it is to identify the source the more (less) equal is the distribution in a normative sense. Sections 4 and 5 follow through two tracks of this framing—the informational surprise when the source of the dollar is revealed, and the number of questions needed to find the source. Section 4 is closest to Theil's (1967) derivation of his inequality measure. Section 5, although it leads to the same mathematical measure, is close to the theory of coding and labelling. This latter interpretation is then used in Section 6 to relate inequality measurement to the Rawls (1971) discourse on the veil of ignorance. Section 7 concludes.

2. Daltonian Inequality Measurement

Dalton's classic paper on inequality measurement begins with the lines:

"It is generally agreed that, other things being equal, a considerable reduction in the inequality of incomes found in most modern communities would be desirable. But it is not generally agreed how this inequality should be measured." (Dalton, 1920, p.348)

Dalton famously sets out criteria that a measure of inequality should satisfy, among them:

[1] For example Cowell (2003), dos Santos and Wiener (2019)

"We have, first, what may be called the principle of transfers…[I]f there are only two income-receivers, and a transfer of income takes place from the richer to the poorer, inequality is diminished." (Dalton, 1920, p. 351).[2]

He goes on to test various measures with respect to this criterion and concludes:

"So far, then, as tested by the principle of transfers, the standard deviation, whether absolute or relative, and the mean difference, whether absolute or relative, are good measures; Professor Bowley's quartile measure is a very indifferent measure; the mean deviation, whether absolute or relative, is a bad measure; and Professor Pareto's measure evades judgment." (Dalton, 1920, p. 354).

Dalton's discussion can be formalized, as he did himself, in mathematical notation as follows. Let x = (x_1, x_2,….., x_n) be a vector of non-negative incomes for i = 1,2,…n individuals. An inequality measure is a function F(x_1, x_2,….., x_n). What properties should the function F(.) satisfy for it to be classed as a "good" inequality measure? The principle of transfers imposes curvature restrictions on F(.), in particular that it have a certain concavity in its functional form. Further properties are also considered in the literature. For example, scale independence imposes the mathematical restriction that F(.) be homogenous of degree zero.

A key property of inequality measures, investigated six decades after Dalton, is that of sub-group decomposability (Cowell, 1980; Shorrocks, 1980). Suppose the population is divided into mutually exclusive and exhaustive groups. Calculate the mean for each group and the inequality for each group. Can overall inequality be written as a function only of group means and inequalities? If so, the measure satisfies a certain type of decomposability. Further, if overall inequality can be written as the sum of two components, one based solely on group means and the other a weighted sum of group inequalities, then the measure is said to be additively decomposable. These restrictions induce a functional equation which can be solved to give the functional form of F(.). These decomposable inequality measures are very important in empirical analysis as they provide useful descriptors of sectoral patterns of inequality.

Dalton (1920) also introduced a line of reasoning which related inequality directly to a social welfare function:

"The objection to great inequality of incomes is the resulting loss of potential economic welfare. Let us assume, as is reasonable in a preliminary discussion, that the economic welfare of different persons is additive, that the relation of income to economic welfare is the same for all members of the community, and that, for each individual, marginal economic welfare diminishes as income increases. Then, if a given income is to be distributed among a given number of persons, it is evident that economic welfare will be a maximum, when all incomes are equal. It follows that the inequality of any given distribution may conveniently be defined as the ratio of the total economic welfare attainable under an equal distribution to the total economic welfare attained under the given distribution. This ratio is equal to unity for an equal distribution, and is greater than unity for all unequal distributions. It may, therefore, be preferred to define inequality as this ratio minus unity, but for comparative purposes this modification of the definition is unnecessary. Inequality, however, though it may be *defined* in terms of economic welfare, must be measured in terms of *income*."

[2] Dalton also gives credit to Pigou in a footnote, referring to Pigou's book *Wealth and Welfare* p.24, thereby leading to our label-- the Pigou-Dalton principle of transfers.

This reasoning reached its culmination half a century later in another classic paper, by Atkinson (1970), where an operational measure was developed relating a social welfare function to inequality. The Atkinson "Equally Distributed Equivalent" measure of inequality allows differing value judgements on inequality aversion to generate inequality comparisons, and is now a workhorse measure in the empirical literature.[3]

The Daltonian perspective is a distinctive approach which can be said to dominate the current inequality measurement literature. But there are alternative approaches and perspectives which can also shed light on how we should conceptualize inequality and its measurement. One such approach is that of information theory, to which this paper is primarily directed.

3. Who Did That Dollar Come From?

Let the total income in the vector x = (x_1, x_2,….., x_n) be X and let mean income be given by µ. Let the share of each individual in total income be denoted by the vector p = (p_1, p_2,….., p_n). Let us now play the following party game. The quiz master holds up a dollar drawn at random from this distribution. The quiz master knows which individual the dollar came from. We do not. Consider now the following two tracks of the game.

Track 1. The quiz master tells us which individual i = 1, 2, …., n the dollar came from. How surprised are we given our knowledge of the vector p? Or, as Shannon (1948) might say, what is the information content of the revelation that an event i has occurred, given the probability distribution of events? As Shannon famously showed, a set of intuitive axioms pin down the information content of event i occurring to be -$\log_2(p_i)$. Thus the expected information content of the whole distribution is $\mathrm{H} = -\sum_i p_i \log p_i$, in other words, the entropy of the distribution.

Track 2. The quiz master asks us to find out who the dollar came from, with the help of a search machine which will give truthful yes/no answers to questions on whether the dollar belongs to a specified subset of individuals. Or, again as Shannon (1948) might say, your task is to devise a sequence of questions which will trace the dollar back to the owner in the fewest possible questions. Equivalently, the task is to design the shortest possible binary code system which distinguishes the messages i = 1, 2, ……n. Shannon's (1948) remarkable theorem is at the smallest number of questions in expectation, or the smallest expected length of the binary code, is $\mathrm{H} = -\sum_k p_i \log p_i$, in other words, the entropy of the distribution.

From an information theory perspective, the two tracks are actually two sides of the same coin, but they lead to interesting and distinctive interpretations of inequality measures, which we will now go on to explore.

4. Priors, Posteriors and Inequality

[3] The Dalton and Atkinson measures are "total" measures of inequality in that they are based on the social welfare gain from total elimination of inequality. Stiglitz (1976) develops "marginal" measures of inequality based on the social welfare gain from a small reduction in inequality.

Shannon's axioms specify that the information content of event i with probability p_i is ($-\log p_i$). Suppose we receive a message saying that the probability is not p_i but q_i. The information value of this message is then

$$-\log(p_i) - [-\log(q_i)] = \log\left({q_i}/{p_i}\right)$$

Suppose the message now transforms each of the i probabilities. For each event there is an informational change. The overall change, in expectation, taken with respect to the new probabilities, is

$$\sum_k q_i \log(q_i/\,p_i\,) = D(q\,||\,p\,)$$

where the expression D is recognized as the Kullback-Leibler divergence measure between the prior distribution p and the posterior distribution q. It is also known as the relative entropy measure. This is a key concept in information theory.

Suppose we start off with the prior belief that the income distribution is perfectly equal, in other words that $p_i = 1/n$ for all i. But we are then told that the actual distribution is

$$q = (q_1, q_2, \ldots, q_n) \text{ with } q_i = x_i/X = x_i/n\mu$$

What is the surprise caused by this new information? Applying these values to D we get:

$$D(q\,||\,p\,) = \sum_i q_i \log(q_i/\,p_i\,) = (1/n)\,\Sigma(x_i/\mu)\log(x_i/\mu) = T$$

The final expression T will be recognized as the famous Theil index of inequality. In fact Theil derived his index following a different route, by noting that when the distribution is equal then entropy is at a maximum, and using the difference between the maximum entropy and actual entropy as a measure of inequality, giving the same expression as above.

No sooner had Theil (1967) arrived at his measure via the Shannon-information-inequality route, he turned "Daltonian":

"To verify whether it is an acceptable measure for inequality we shall apply the following…. tests…." (Theil, 1967, p. 92).

He showed that it satisfied scale independence and the principle of transfers, and he celebrated the fact that it was decomposable in a particular sense:

"Our starting point was the entropy H as a measure of equality, which we modified slightly to an inequality measure by subtracting H from its maximum value. We now find that this measure has a simple interpretation in terms of income shares and population shares: moreover, that it can be aggregated in a straightforward manner. In that respect it is more attractive than most well-known inequality measures such as Gini's concentration ratio" (pp 95-96).

It is striking that no further use is made of the information theoretic foundations in Theil's subsequent discussion of the normative properties of his measure. But the priors versus posteriors perspective provides a useful conceptualization of inequality—it is the surprise, or normative jolt, received by a going-in belief in equality when confronted by a world that is not equal. The formula is the same, but the interpretation is different and provides new insight and intuition.

An application from Kanbur and Snell (2019) will illustrate this point. Start again with the notion of surprise at an actual unequal distribution when the prior is that of an equal distribution. This perspective can be used to develop a test for "fairness" as follows. Imagine a helicopter which has a basket full of X dollars to drop, one by one, on to a population of n individuals. Let the probability that a dollar sticks to individual i be p_i. Then p_i= 1/n is a specification of "equality of opportunity" in this world where no other attributes of individuals are specified.

After the helicopter drop, we observe an actual distribution of dollars across individuals:

$q_i;\ i = 1,\ 2,\ \ldots,\ n$

On the basis of these observations we would like to test for the hypothesis that the process was fair, in other words, that

p_i= 1/n for all i.

Consider the Likelihood Ratio test for this null hypothesis. It is shown in Kanbur and Snell (2019) that the LR test statistic for this hypothesis under the multinomial process of the helicopter drop is in fact proportional to the Theil Index of Inequality:

$$LR \propto T$$

The asymptotic distribution of this test statistic is known and can be implemented (Kanbur and Snell, 2019).

The central take away is that, going beyond the standard Daltonian and Atkinsonian perspectives, the Theil Index of Inequality can be interpreted as a test statistic for the hypothesis of fairness, for a specified income generating process and a specified notion of fairness within that process. Inequality indices, in this perspective, are no longer just ex post measures of welfare loss from inequality. They can be turned to use as test statistics for fairness.

Posing the issue of inequality measurement as one of divergence between equality as a prior and inequality as the posterior in an information theoretic framework leads to the Theil index. What if we were to keep the information theoretic framework but reverse the sequence and think of inequality measurement as quantifying the divergence between inequality as the prior and equality as the posterior? The Kullback-Leibler divergence is then given by:

$D(p\,||\,q) = \sum_i p_i \log(p_i/\,q_i\,) = (1/n)\,\Sigma\log\,(\mu/x_i) = L$

This is just the Mean Log Deviation (MLD).

L is also known as "Theil's second measure" because, although it also satisfies the properties he set out as desirable, its derivation was relegated to an Appendix to Chapter 4 on The Measurement of Inequality. This is partly because he struggled with the sequence going from the actual distribution as the prior to the equal distribution as the posterior:

"….one may argue that it is against intuition not to take the equal population shares as a starting point but to consider them as "posterior". But this is a matter of taste." (Theil, 1967, p. 127).

Since both T and L satisfy scale independence and the principle of transfers, and are also decomposable, how can we decide which measure is "better"? Theil's argument given above is that it is more intuitive to go from an equal distribution as the prior to the actual distribution as posterior than the other way around, thus favoring T. But Shorrocks (1980) makes a counterargument in favor of L on the basis of the specific way in which T and L are decomposable. He poses the issue in terms of an applied empirical exercise one often conducts:

"When inequality measures are used to assess the contribution of one particular factor to total inequality, [a] problem arises in the different interpretations that can be placed on statements like 'X per cent of inequality is due to Y." Consider, for example, the question "How much inequality can be attributed to age variations in income." This may be interpreted as meaning: (i) How much less inequality would we observe if age variations were the only source of income differences; (ii) by how much would inequality fall if age-income differences were eliminated….Interpretation (i) suggests comparison of total inequality with the amount which would arise if inequality was zero within each age group, but the difference in mean income between age groups remained the same….Interpretation (ii) suggests a comparison of total inequality with the inequality value which would result if the mean incomes of the age groups were made identical, but inequality within each age group remained unchanged." (p. 624).

With this framing Shororocks (1980) then sets out a further criterion for a "good" inequality measure going beyond additive decomposability in a general sense. This is that (i) and (ii) should give the same answer. This criterion is only satisfied by L:

"For this reason, [L] is the most satisfactory of the decomposable measures, allowing total inequality to be unambiguously split into the contribution due to differences between sub-groups." (p. 625).

The specifics of decomposability may be the right way to choose between inequality measures in the Daltonian spirit. But an alternative is to directly engage with the prior-posterior framing from information theory and choose the sequence which most accords with our intuition. In any event, if we choose L as the index then it is consistent with a normative perspective which sees inequality as the divergence between the actual distribution as the prior and the equal distribution as the posterior.

Finally, consider departures from specifying the alternative to the actual distribution, whether as prior or posterior, not as the equal distribution but as another norm vector derived from other principles. Why should the norm be restricted only to perfect equality? Other considerations such as equality of opportunity could be introduced, as is done by Roemer (1998). Combinations of equality of opportunity and poverty focused norms can also be considered. Such elaborations are provided for example in Hufe, Kanbur and Peichl (2022) and lead to alternative measures which can better measure the unfair component of inequality.

5. Coding and Questions

Recall from Section 3 the random draw of a dollar from a distribution of income across n individuals. Track 2 asks us to find out who the dollar came from, by asking the smallest number of questions with yes/no answers. The normative intuition is that the more equal is a distribution, the more difficult it would be to find out who the dollar came from. This problem from income distribution is in fact isomorphic to a problem in binary coding addressed by Shannon (1948).

Consider the following example.[4] Let there be five letters A, B, C, D, E which appear in a language, with counts 3, 5, 4, 6, 2 respectively and thus relative frequencies 3/20, 5/20, 4/20, 6/20, 2/20. Our task is to design binary coding which distinguishes the letters and which has the shortest possible overall code length. Such a data compression exercise was posed in general form by Shannon (1948) in his classic paper. An answer to the question was provided by Huffman (1952) with a method which is now known as the Huffman code or Huffman algorithm. First rank the letters from smallest to largest frequency. Then combine the two smallest frequency into one. Treating this as a new frequency, combine the two smallest frequencies in the new list into one. Continue until, of course, all letters are combined into one unit. Represent this process as a tree with nodes and branches and label the branches coming out of each node as 0 on one side and 1 on the other. For the specific example taken here, the tree is given in the Figure below.

The tree and the labels on its branches give the Huffman binary code for the five letters in the example:

E: 000
A: 001
D: 01
B: 10
C: 11

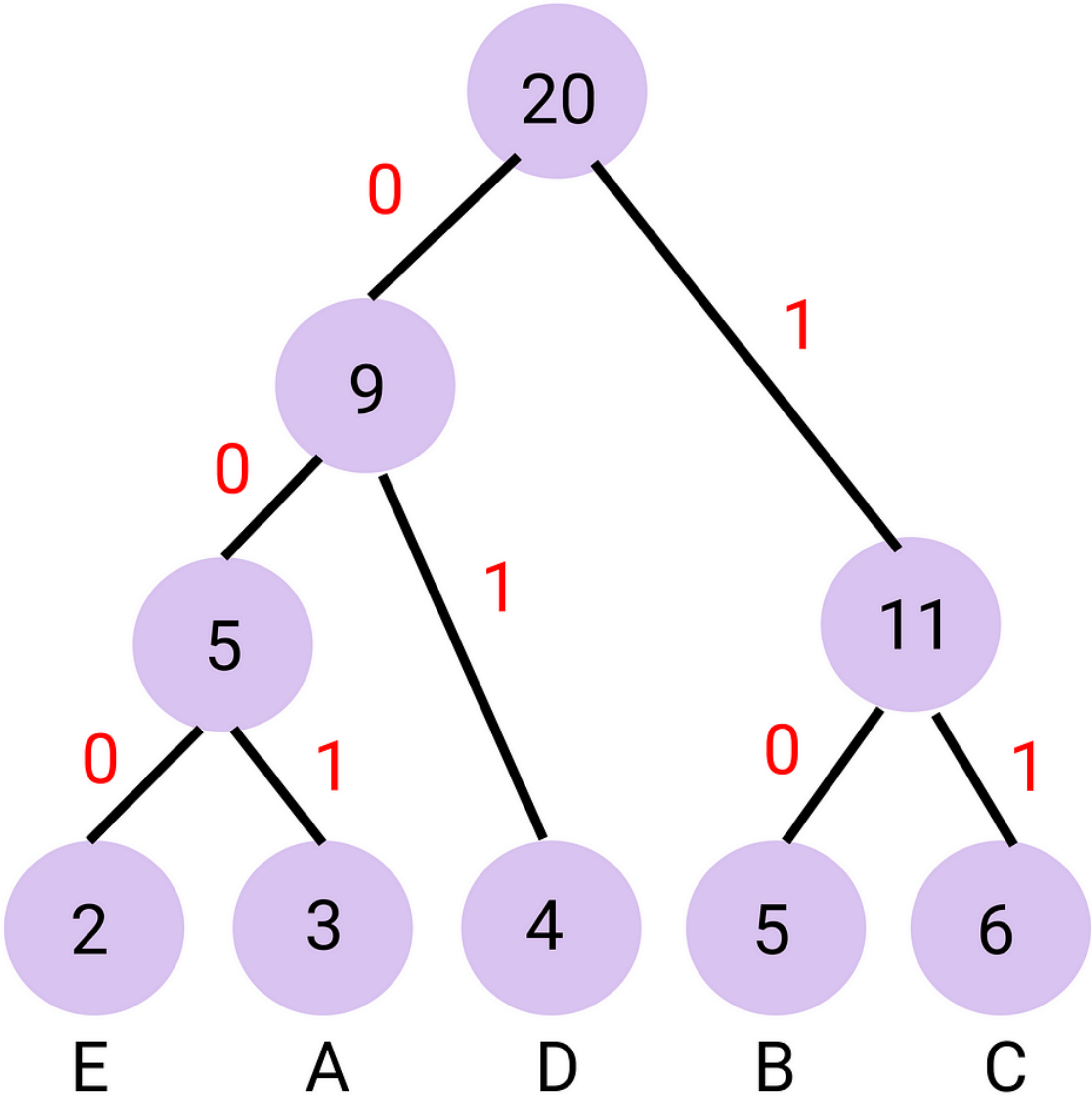
The procedure is of course quite general and can be applied to any distribution p = (p_1, p_2,….., p_n). The Huffman code minimizes the expected code length within a particular class of binary codes, but it was shown by Shannon (1948) in a remarkable theorem that the smallest possible expected code length within a general class of codes is in fact

$$\mathrm{H} = -\sum_k p_i \log p_i,$$

in other words, the entropy of the distribution. The formula also shows that the optimal coding structure has code length $(-\log p_i)$ for the i^{th} letter or entity.

Note that the structure of the tree also gives the structure of the questions which will find for us "who the dollar came from." Looking at the tree, we would first ask "was it from B or C". If the answer is yes we would ask "was it B". If the answer is no we would ask "was it from D". If no we would ask "was it from E or A". The point is that this structure of questioning is the one which gets us the answer in the smallest number of questions in expectation. Optimal coding and the optimal sequence of questioning are embedded in each other, as is seen from the tree.

[4] The example and the Figure shown are taken from Saglam (2020) https://medium.com/makepad/huffman-coding-compression-algorithm-d68b098f768b


20
0
1
9
0
1
5
11
0
1
0
1
2
3
4
5
6
E
A
D
B
C

Thus for any distribution p = (p_1, p_2,….., p_n) there exists an optimal coding structure with code length $(-\log\ p_i)$ for each p_i. Similarly, for any distribution q = (q_1, q_2,….., q_n) there exists an optimal coding structure with code length $(-\log\ q_i)$ for each q_i. But suppose we were to use the wrong code for the wrong distribution. If we take the optimal coding scheme for p and apply it to q, then the expected code length would be

$$-\sum_i q_i \log\ p_i$$

This must be longer than the optimal code length for q, where the code length for each i would be $(-\log\ q_i)$. The cost of using the wrong code for q, measured by the increase in the expected code length as result, is given by

$(-\ \sum_i\ q_i \log\ p_i)\ -(-\ \sum_i\ q_i \log\ q_i)$ = $\sum_i q_i \log(q_i/\,p_i\,)$

= $D(q\,||\,p\,)$

which of course is the Kullback-Leibler measure of divergence between p and q.

We saw in section 5 that the Kullback-Leibler measure of divergence between p as the equal distribution and q as the actual distribution is in fact the Theil index of inequality T. We now have an alternative interpretation of T. It is the cost, in terms of increase in expected code length, of applying the optimal binary code for the equal distribution p but applying it to the distribution q. It is also the cost, in terms of increase in the number of expected questions, of taking the optimal schema of questions for finding out who the dollar came from when the distribution is equal, but applying it to the actual distribution at hand. It should also then be clear that on the other hand if we were to ask the cost of applying the code for the actual distribution to the equal distribution, the answer would be L, the Mean Log Deviation. The next section turns to an interpretation of these relationships in terms of the Rawlsian Veil of Ignorance.

6. Codes, Labels and the Veil of Ignorance

A foundational construct of the modern discourse on justice and inequality is Rawls's (1971) "original position" and the concomitant "veil of ignorance":

"Among the essential features of this situation is that no one knows his place in society, his class position or social status, nor does anyone know his fortune in the distribution of natural assets and abilities, his intelligence, strength, and the like. I shall even assume that the parties do not know their conceptions of the good or their special psychological propensities. The principles of justice are chosen behind a veil of ignorance. This ensures that no one is advantaged or disadvantaged in the choice of principles by the outcome of natural chance or the contingency of social circumstances." (Rawls, 1971, p. 12)

It is from the original position behind the veil of ignorance that Rawls draws his famous twin principles of justice, of which the second one, the "difference principle" that inequalities are justified only in so far as they benefit the worst off, the "maximin", is now a basic component of economic analysis.

There has been a mountain of writing on the veil of ignorance which would be impossible to provide commentary on here. But one major strand of debate is on the "realism" of the original position:

"Many criticisms have been leveled against Rawls's veil of ignorance. Among the most frequent is that choice in the original position is indeterminate (Sen, 2009, 11–12, 56–58). Among other reasons for this, it is said that the parties are deprived of so much information about themselves that they are psychologically incapable of making a choice, or they are incapable of making a rational choice" (Freeman, 2023)

Despite Rawls's (1971) defense that

"This original position is not, of course, thought of as an actual historical state of affairs, much less as a primitive condition of culture. It is understood as a purely hypothetical situation characterized so as to lead to a certain conception of justice" (p. 12)

and despite his adducing the authority of Kant himself to justify this perspective,[5] the debate continues.

The intuitive power of Rawls's attempt to "wipe out" actual circumstances including income of each individual when inviting them to conceptualize justice and derive the principles of justice, comes up against the equally intuitive perception that it is in practice difficult to achieve this even as a thought experiment for individuals living in a world that is highly unequal in actuality. But could we quantify how difficult it is for such "wiping out" to happen? Codes and labels could provide us one route to answering this question.

The previous section has derived efficient coding for elements in a probability distribution, be it a distribution of frequencies across letters of the alphabet, or a distribution of income across individuals. It has referred to Shannon's remarkable result that the minimum code length, in expectation, for a code which uniquely distinguishes individuals (or letters) is none other than the entropy of the distribution. We further considered the possibility of applying the wrong code to the wrong distribution and thereby paying a cost in term of increased expected code length.

Consider now the Shannon binary code for each individual as being a label which distinguishes each individual from others in terms of income. It conveys information on income efficiently (in the sense of minimum expected code length). It could be argued that the most efficient method is surely to simply have a sign on each individual giving the number of dollars. But we do not do that in society. Rather, we have indirect correlates which signal income and status. These might include accent, clothes, other acquaintances or college (these last two to be dropped causally in conversation). Intuitively and anecdotally, these signals or labels are quite powerful in distinguishing individuals and in signaling their income social status. It is these labels, efficient as they are in separating individuals from each other, that need to be "wiped out" to get us behind the veil of ignorance.

The existing labels are wiped out. But what is put in their place? "Nothing" does not seem right, since these are then not individuals capable of making rational choice, in line with critiques referred to above. Consider the replacement candidates as the set of labels that would apply for a perfectly equal society where income would not be a differentiator between individuals. This is clearly an effort of will

[5] Footnote 5 on p. 12 of Rawls 1971) says: "Kant is clear that the original position is hypothetical. See *The Metaphysics of Morals,* pt. I *(Rechtslehre),* especially pp 47, 52…."

for an actually unequal society. One measure of the cost of this is the increase in expected code length when the codes (labels) for an equal society are used in the current unequal society. But, as shown in the previous section, this is

$$(-\sum_i q_i \log p_i) - (-\sum_i q_i \log q_i) = \sum_i q_i \log(q_i/p_i)$$

$$= D(q \,||\, p)$$

which is the Kullback-Leibler divergence between the actual distribution q and the veil of ignorance distribution p. When p is the equal distribution ie $p_i = 1/n$ for all i, then this is simply the Theil index of inequality T.

This line of reasoning then provides yet another interpretation of the Theil index of inequality beyond its Daltonian properties and as a test statistic for fairness as developed in the previous sections. It now appears as a quantitative measure of the difficulty of implementing the veil of ignorance, seen as the cost of wiping out individual labels induced by the current unequal distribution to one that would pertain if the distribution were equal.

7. **Conclusion**

This paper has shown the power of applying Shannon's (1948) information theory perspective to inequality measurement by considering the thought experiment of drawing a dollar at random from an income distribution and asking who the dollar came from. The surprise at being told who the dollar came from, and the task of designing a set of questions with yes/no answers which will get us to the person, are two sides of the same coin but with interesting interpretations. The Theil index of inequality, which Theil (1967) himself derived with reference to information theory and entropy but did not then explore further is shown to have interpretations beyond its simple Daltonian properties such as satisfying the principle of transfers or being sub-group decomposable. It can be interpreted as a statistical test of the hypothesis of fairness, and as a quantitative measure of the difficulty of achieving Rawls's (1971) original position behind the veil of ignorance.[6]

[6] In a critique of the Theil index, Sen (1997, p. 36) says "But the fact remains that [the Theil index] is an arbitrary formula, and the average of the logarithms of the reciprocals of income shares weighted by income is not a measure that is exactly overflowing with intuitive sense." The arguments and interpretations provided in this paper perhaps go some way towards redressing this sentiment.

References

Atkinson, A. B. 1970. “On the Measurement of Inequality,” *Journal of Economic Theory,* vol. 2, pp. 244-263.

Cowell, F. A. 1980. “Generalized entropy and the measurement of distributional change”. *European Economic Review* 13, 147—159.

Cowell, F.A. 2003. “Theil, Inequality and the Structure of Income Distribution.” Discussion Paper DARP 67, London School of Economics.

Dalton, Hugh. 1920. “The Measurement of the Inequality of Incomes”, *Economic Journal*, Vol. 30, No. 119, pp. 348-361.

Dos Dantos, Paulo, and Noe Wiener. 2020. “By the content of their character? Discrimination, social identity, and observed distributions of income.” *The Journal of Mathematical Sociology*, Volume 44.

Freeman, Samuel. 2023. "Original Position", *The Stanford Encyclopedia of Philosophy* (Winter 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/win2023/entries/original-position/>.

Hufe, Paul, Ravi Kanbur and Andreas Peichl. 2022. “Measuring Unfair Inequality: Reconciling Equality of Opportunity and Freedom from Poverty.” *Review of Economic Studies*, Vol. 89, Issue 6, pp. 3345-3380.

Huffman, David. 1952. “A Method for the Construction of Minimum-Redundancy Codes.” Volume 40, Issue 9, pp 1098-1101.

Kanbur, Ravi and Andy Snell. 2019. “Inequality Indices as Tests of Fairness” *Economic Journal*, https://doi.org/10.1111/ecoj.12637

Rawls, John. 1971. *A Theory of Justice*. Harvard University Press.

Roemer, John. 1998. *Equality of Opportunity*. Harvard University Press.

Saglam, Idil. 2020. “Huffman Coding Compression Algorithm”. https://medium.com/makepad/huffman-coding-compression-algorithm-d68b098f768b

Sen, A. (1997). *On Economic Inequality*. Oxford: Clarendon Press.

Shannon, C. E. 1948. “A Mathematical Theory of Communication.” *Bell System Technical Journal* 27: 370-423.

Shorrocks, Anthony F. 1980. “The Class of Additively Decomposable Inequality Measures.” *Econometrica*, Vol. 48, No. 3, pp. 613-625.

Stiglitz, Joseph E., 1976. “Simple Formulae for the Measurement of Inequality and the Optimal Linear Income Tax,” IMSSS Technical Report No. 215, Stanford University, August.

Theil, Henry. 1967. *Economics and Information Theory*. North Holland.