The Informational Index of Income Inequality

Paulo L. dos Santos*

July 27, 2022

Abstract

A new informational index of income inequality is proposed. Based on a combinatorial characterisation of entropy and on its generalisation by multivariate measures of co-information, the index addresses a limiting analytical choice embedded in Theil's two indices of income inequality. This yields a positive measure of inequality of opportunity in income-generating processes, defined in relation to probabilities that the population-wide distribution of income describes the set of possible income levels facing specified sub-groups of that population. Its measure across a population is given when sub-groups consist of each individual. This can be successively decomposed linearly by sub-groups defined by covariates of income. In those instances, the index also provides informational measures of phenomenological association and interaction between income and those covariates. On those bases the index improves on mean-log deviations as measures of "luck egalitarian" notions of equity; casts new light on and uncovers new properties in the Theil-Finezza index of segregation; offers a non-parametric generalisation of the Kitagawa-Oaxaca-Blinder decomposition; and lays new conceptual foundations for work on the determinants and normative content of patterns of income differentiation in decentralised market economies.

Keywords: Inequality Indices, Income Distribution, Information Theory
JEL codes: C18, D32, D63

1 Information Theory and Analysis of Inequality

In chapter four of his seminal *Economics and Information Theory*, Henri Theil, 1971 developed and established the key mathematical properties of two entropic indices of income inequality. While those indices have been widely used as measures of inequality over the past fifty years, considerable controversy remains about their conceptual foundation. Amartya Sen, 1997 expressed a widely shared concern when noting that Theil's first 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."

*I would like to thank Ravi Kanbur, Duncan Foley, Sanjay Reddy, Amos Golan, Noé Wiener, Ellis Scharfenaker, and Jenn Yablonski for the many exchanges that helped me arrive at what is reported here. The usual disclaimers apply.

The controversy is understandable. For many decades after the introduction of the concept of entropy, disagreement persisted among physicists on why exactly it helped them establish some of the most empirically successful theories in physics.[1] Some of the resulting conceptual ambiguities were spread into economics by a number of interventions that have sought to import the concept of entropy into economic analysis somewhat mechanically, without explicitly addressing exactly why entropy functionals may be useful in analysis of economic systems. Matters were not helped by Theil himself. His conceptual exposition was arguably terse, motivating his indices because they offer measures of inequality that obey the Dalton-Pigou Transfer Principle and are linearly decomposable across sub-groups, without a discussion on what entropic functionals may measure for distributions of an economic variable like individual income.

Two recent contributions have effectively underlined the importance of achieving conceptual clarity concerning the use of entropic functionals in analysis of income distributions. Each has independently considered the possibility that entropic functionals of income distributions may be be understood as more than just descriptive, and may be related to characterisations and to normative assessments of the processes shaping income distributions. Kanbur and Snell, 2019 show how within the terms of well-specified "helicopter drop" models of income-generating processes, Theil's indices can be understood as test statistics for the presence of Friedman, 1962's notion of fairness or *ex ante equality of treatment*. Taking a different approach to the use of entropic functionals in economic analysis, dos Santos and Wiener, 2020 showed how entropic measures of joint distributions of income and its covariates offer robust, informational measures of association between income and its covariates, offering a powerful new way to grapple with the role played in income-generating processes by what Roemer, 1998 and other "luck egalitarians" term "circumstantial" individual characteristics.

Those contributions raise the prospect of drawing on the analytical power of information theory to support work on the conceptualisation, measurement, and determinants of income inequality.[2] They suggest that entropic functionals offer bases for indices of income inequality with specific phenomenological content, exhibiting a clear probabilistic relationship to aspects of income-generating processes. Pursuing these intriguing analytical possibilities requires a clear understanding of the bases for using information-theoretic measures in analysis of complex processes in economic systems, like those generating levels of individual income. Unfortunately, there is no widespread, shared understanding among either economists or physicists interested in economic processes of what those bases may be.

This paper makes three related contributions in this connection.

[1] See Jaynes, 1989, for a discussion of the differences between appreciations of the Principle of Maximum Entropy based on the "ergodic principle," and his own epistemic or combinatorial interpretation, which he traces back to the work of Willard Gibbs, 1902 and Claude Shannon, 1948, and guides the approach taken by this paper.

[2] Jaynes, 2003 and Csiszar, 1998 offer comprehensive and authoritative discussions on the general applicability of information-theoretic concepts to probabilistic analysis of a broad range of systems.

First, it offers a detailed conceptual discussion of how entropic functionals can be used to guide quantitative inquiry into the functioning of any system composed of large numbers of distinguishable individual members that, at any given point in time, occupy specific individual states.[3] This includes economic systems. In grappling with such systems, inquiry is often concerned with the nature of and relationships between *micro-level configurations* and *macroscopic states* involved in their functioning. The former are descriptions of the arrangement of each of the system's members across all possible individual states in which they may find themselves. The latter are distributions offering aggregate descriptions of the frequency with which each possible individual state is occupied across the system. Entropy is a combinatorial, multinomial measure of the relationship between micro- and macro-level states, defined for macroscopic states or frequencies. It quantifies the *multiplicity* or number of micro-level configurations that result in a given macroscopic state or frequency. This simple concept is eminently useful in thinking about issues of aggregation in complex economic systems,[4] and in statistical inquiry into their functioning.

Second, clarity on the combinatorial content of entropy functionals highlights a key analytical limitation in Theil's adoption of those functionals to measure income inequality. Use of Theil's indices to grapple with the nature of income-generating processes creates the same type of difficulty first identified by Bose, 1924 and Einstein, 1925 in the application of conventional multinomial statistics to analysis of aggregations of Bosons—elementary particles that are indistinguishable when occupying the same energy state. That is because in his indices of income inequality, Theil effectively treated "units of income" or "units of money" as the basic unit of analysis or members of the economic system, leaving individual agents as "states" those members may occupy. But money balances are perfectly fungible, ensuring "units of income" or money are economically indistinguishable. All micro-configurations of "units of money" across individuals giving rise to the same macro-level distribution are in fact the same economic micro-state. As a result, the multinomial functionals at the heart of Theil's indices do not provide the correct measure of the micro-level multiplicity of any given distribution of "units of money" across individuals. While not changing the ability of Thiel's indices to offer descriptive measures of income inequality, this limits their broader analytical usefulness a great deal.

Third, drawing on the combinatorial interpretation of entropy and on its multivariate generalisation by measures of *co-information*, the paper develops a new *informational index of inequality*. When applied to individual incomes, the new index avoids the limitations present in Theil's indices of income inequality, unlocking the full analytical power of entropic functionals.

The index offers a well-specified, probabilistic measure of *inequality of opportunity*. The measure is positive, in that it requires neither specification of what may or may not be a "fair" income

[3] This approach was first articulated in economic analysis by Foley, 1994. See also its development in Caticha and Golan, 2014 and dos Santos, 2020. The latter contribution has termed it the socio-combinatorial or Jaynes-Gibbs application of entropic functionals to quantitative analysis of economic systems.

[4] As advocated as early as Stutzer, 1983.

distribution, nor any normative characterisation of social welfare. Instead, it is defined in relation to probabilities that specified sub-groups in a population of income earners faced the same set of possible income levels in income-generating processes. When sub-groups consist of each individual in a population, the index offers a negative, logarithmic measure of the probability that each and every individual's income was the result of independent drawings from that population-wide distribution. When sub-groups are defined by covariates of income, the index offers an equivalent probabilistic measure of inequality of opportunity across those groups. The index also offers very general measures of phenomenological association. Its measures across subgroups defined by any covariate of income correspond to well-defined, informational measures of association and interactions between income and that covariate in income-generating processes. Its value across a population offers an informational measure of the significance of individual identity in those processes.

The index defines structures of successive, linearly additive decompositions of its measure of inequality in a marginal distribution of income. All elements in those decompositions have well-defined content as probabilistic measures of inequality of opportunity, and as informational measures of association and interaction between income and the covariates defining them. As such, the index not only exhibits a general, multivariate instance of the decomposability considered by Shorrocks, 1980 and by the contributions summarised in Cowell, 2006, but it also defines successive measures of across- and within-group inequality that correspond to the structures of phenomenological association between income and the covariates defining the decompositions in question.

By offering coincident, informational measures of inequality and association, the index defines a number of new results and new avenues of inquiry. It defines a non-parametric, informational generalisation of the Kitagawa-Oaxaca-Blinder decomposition of the total influence any given set of characteristics exerts over individual incomes.[5] It helps identify new properties in the measure of segregation proposed by Theil and Finezza, 1971, including its interpretation as a test statistic. It also offers distinctively useful tools for normative inquiry on the determinants of income inequality, including robust and very general way to instrumentalise "luck egalitarian" notions of distributive justice put forward by Roemer, 1998 and others.

The paper is organised as follows. Section two briefly discusses Theil's two indices of income inequality and their salient mathematical properties, and discusses recent uses of entropic measures in work on income inequality. Section three offers a necessary and unavoidably technical discussion concerning the application of information theory to analysis of any system made up of a large number of members. That includes an explanation of entropy functionals as a combinatorial heuristic that can help guide inquiry into the functioning of those systems, as well as an informational measure of observer uncertainty, heterogeneity, and diversity implicit in a system's macroscopic state. Section four draws on those explanations to show how Theil's indices of income inequality embody a specific analytical choice that greatly limits their analytical usefulness. Sections five and

[5] Kitagawa, 1955; Oaxaca, 1973; Blinder, 1973.

six define the two central building blocks of the informational index of income inequality–the mutual information and its multivariate generalisation by measures of co-information. The discussion draws on recent advances in information theory to characterise the phenomenological content of informational interactions. It also identifies, possibly for the first time, their content as probabilistic measures of the presence of a specific symmetry in the functioning of any system—a symmetry that may be naturally associated with notions of inequality of opportunity in social inquiry.

This sets the stage for the definition of the informational index of income inequality and a discussion of its salient mathematical properties in section seven, followed by a concluding discussion of how the index usefully helps re-frame conceptual and practical inquiry on the factors and associations driving differentiations by individual incomes in decentralised market economies.

2 Theil's Indices, Divergences, and Income-Generating Processes

Henri Theil put forward two entropic indices of income inequality. In Theil, 1971, those are considered for distributions of income across a population of $m = \{1, 2, ..., N\}$ agents. Letting the income of each of those individuals have a measure y_m relative to total income, $\mathbf{y}$ represent vectors of those values across the population, and ρ_m denote each individual's income relative to average income, the two indices are given by,

$$T_1(\mathbf{y}) = \sum_m y_m \log y_m + \log N \; ; \qquad T_2(\mathbf{y}) = -\frac{1}{N}\sum_m \log \rho_m \tag{1}$$

As is well known, these indices of inequality are additively decomposable, obey the Dalton-Pigou principle of transfers, and do not change under proportional changes across all incomes.

2.1 Theil's Indices as Directed Divergences

Theil's two indices are also understood as formal measures of the difference between a distribution of income $\mathbf{y}$ and a perfectly equal distribution of income across individuals. This difference is given by the Kullback-Liebler divergence, a directed informational measure of the extent to which a distribution $\mathbf{p}$ differs from a reference distribution $\mathbf{q}$, formally given by,

$$D(\mathbf{p}, \mathbf{q}) = \sum_m p_m \log \frac{p_m}{q_m} \tag{2}$$

Thiel's two indices are in fact conjugate measures of that divergence. This can be expressed in at least two useful ways. First, the indices can be understood as measures of the divergence between a distribution of income $\mathbf{y}$ and a uniform distribution $\mathbf{u}$. Under a uniform distribution, each individual receives the same share $u = N^{-1}$ of total income, ensuring that, for the first index,

$$T_1(\mathbf{y}) = \sum_m y_m \log \frac{y_m}{u} = D(\mathbf{y}, \mathbf{u}) \tag{3}$$

Along the same lines, the second index can be expressed as,

$$T_2(\mathbf{y}) = -\sum_m u \log \rho_m = \sum_m u \log \frac{u}{y_m} = D(\mathbf{u}, \mathbf{y}) \tag{4}$$

Both indices effectively measure the same logarithmic differences between each individual's relative income under distribution $\mathbf{y}$ and what their income would be under a perfectly even distribution $\mathbf{u}$. The first index weighs those differences by the relative income received by each individual, while the second one weighs all individuals equally. In both cases, the indices are ultimately grounded on a specific notion of equality, defined by an equal distribution of income across all individuals.

A second pair of conjugate expressions of Theil's indices as divergences casts additional light on their content. Consider a grouping scheme of the N individuals into groups g containing n_g individuals, each with the same income share y_g. Under those groupings,

$$T_1^g(\mathbf{y}) = \sum_g n_g y_g \log N y_g = \sum_g c_g \log \frac{c_g}{f_g} = D(\mathbf{c}, \mathbf{f}) \tag{5}$$

Where $c_g = n_g y_g$ denotes the share in total income of group g, and $f_g = \frac{n_g}{N}$ is the share of the total population in that group.

Along the same lines, the second index can also be expressed for this kind of grouping,

$$T_2^g(\mathbf{y}) = -\sum_g n_g \log N y_g = \sum_g f_g \log \frac{f_g}{c_g} = D(\mathbf{f}, \mathbf{c}) \tag{6}$$

Here Theil's two indices of income inequality appear as measures of the divergence between the distributions of people and money across groups of people having the same levels of income. Each index emphasises a different aspect of income inequality. In expression 5, logarithmic differences between the two distributions are weighted by the share of total income in the groups in question; in expression 6, the same differences are weighted by the number of people in the group.[6]

Decomposability highlights an important difficulty with Theil's indices. As first identified in Shorrocks, 1980, it is not at all clear how attributions of income inequality to variations across groups defined by any given characteristic may be related to the influence exerted by that characteristic on income-generating processes. In fact, Shorrocks noted that by weighing logarithmic

[6]Equations 5 and 6 offer a formal, information-theoretic statement of a characterisation first raised rather cryptically by Theil, 1971 and emphasised more recently and explicitly by Conceicao et al., 2000.

differences by income shares, Theil's first index is compatible with two distinct ways to understand the measure of inequality "due to" a characteristic like age: The measure of inequality that would be removed if all age-related inequality were eliminated, and the inequality that would be left if all inequality due to factors other than age were removed.

While Theil's second index avoids this problem, the broader difficulty remains: Without a clear understanding of what aspects of income-generating processes a measure like the average mean logarithmic deviations captures, it is not at all clear how measures of cross-group inequality given by either of Theil's two indices may be understood as measures of the inequality "due to" the characteristics defining the groups in question. Resolving this difficulty requires a clear understanding of the connection between any measure of income inequality and income-generating processes.

2.2 Entropic Measures, Income-Generating Processes, and Fairness

In this connection, the equivalence between Theil's indices and an information-theoretic measure of differences between distributions raises an intriguing prospect. Those measures have a clear and well-understood probabilistic content that is generally applicable to inquiry into the functioning of any system composed of large numbers of distinguishable individual members.[7] As such, entropic indices of income inequality may be understood as much more than descriptive. They may be related to probabilistic characterizations of the processes shaping distributions of income, and to formal tests of different normative standards of fairness in those processes.

This was raised recently and independently by two sets of contributions. Kanbur and Snell, 2019 discuss how Theil's indices can be used in parametric tests for the presence of the kind of *ex-ante* "equal treatment" emphasised by Friedman, 1962 as the only legitimate criterion of fairness in income-generating processes. They show how the indices can be related to test statistics for this kind of fairness, under two specific "helicopter-drop" models of income-generating processes: One where units of money are individually allocated to individuals with the same probability of receiving each unit, and another where each individual receives a flow of income units across time until they are cut off, with each individual facing the same instantaneous probability that they will be cut off. Based on explicit consideration of the ways in which entropic functionals may be used in analysis of complex economic systems, dos Santos and Wiener, 2019 and dos Santos and Wiener, 2020 proposed the use of those functionals as informational measures of association between individual incomes and other individual economic characteristics. When applied to associations between incomes and what "luck egalitarians" term circumstantial individual characteristics, entropic functionals provide general measures of the influence those characteristics exert over economic outcomes, pointing to the presence of "inequality of opportunity".

The present discussion can be understood as a synthetic, critical development of these contributions. By drawing and building on the deliberate application of information theory to analysis

[7] Jaynes, 2003; Csiszar, 1998.

of economic systems offered by the second set of contributions, it develops an entropic index of income inequality offering a generalised version of the tests of fairness sought by the first one. Doing this and grappling with the index's properties and distinctive analytical usefulness requires deliberate, technical discussion; including an explicit exposition of how entropic measures can be used in analysis of economic systems, to which the paper now turns.

3 Entropy and Economic Analysis

Since Henri Theil's pathbreaking contribution, a number of scholars have pursued applications of concepts and tools from information theory and statistical mechanics to economic analysis.[8] At the heart of many of these contributions is the concept of *entropy*, the foundational concept of information theory. While entropy has several mathematically equivalent but conceptually distinct foundations, it is most generally defined for quantitative analysis of economic and social systems as a combinatorial tool.[9] Understood as such, entropy offers a simple measure of statistical relationships between the micro-level characteristics or states of economic individuals and the macro-level states of economic systems.

This section outlines a socio-combinatorial characterisation of entropic measures that draws on J Willard Gibbs' approach to the canonical ensembles of thermodynamics and the epistemic understanding of probabilities and information theory emphasised by mathematical physicist ET Jaynes (whose antecedents include the views on probability theory of Keynes, 1921). In the resulting framework, entropy appears as a well-defined measure of heterogeneity and diversity in the individual states of members of a system, and—thus—of the uncertainty an observer aware only of a system's macroscopic state has about them. The framework also points to the limiting analytical choices Theil made in the definition of his indices. As the concepts involved are not widely understood amongst economists, a careful and unavoidably technical discussion is necessary to motivate them adequately, and to convey the economic content of entropic functionals.

3.1 Micro-Configurations, Phase Spaces, and Macro-States

The approach Gibbs, 1902 took to his thermodynamic ensembles can be usefully applied to the problems of quantitative analysis of economic and social systems.[10] The approach offers a robust and novel way to think about the relationship between micro- and macro-level functioning in economic systems—a way that avoids the pitfalls of both aggregative models postulated without

[8]See Georgescu-Roegen, 1971; Farjoun and Machover, 1983; Stutzer, 1983; Foley, 1994; Mantegna and Stanley, 2000; Rosser, 2008; Yakovenko, 2007; Golan, 2018; Caticha and Golan, 2014, as well as the contributions discussed in Scharfenaker and Yang, 2020.

[9]Niven, 2007.

[10]See dos Santos, 2020.

much consideration of micro-level functioning, and of naive "representative agent" characterisations of macroscopic relationships.[11]

Formally, consider an economic system composed of a large number of individual members, $n = 1, 2, ..., N$. Suppose the state of each member is described by a set of $v = 1, 2, ..., k$ individual degrees of freedom, $X = \{X_1, ..., X_k\}$. These may in principle represent any relevant characteristic, contemporaneous or past. They may be quantifiable or denote qualitative or categorical individual characteristics, including descriptions of an individual's institutional or relational situations. Let $\mathbf{x_n}$ denote the individual state occupied by member n.

In all practical economic inquiry the domain of each quantitative individual degree of freedom v is "coarse grained" into s_v distinct value ranges or "bins". Qualitative characteristics may also be coarse grained with coding schemes mapping them to distinct real numbers. Coarse graining defines a set of $i = 1, 2, ..., s$ possible individual states $\mathbf{x^i} = \{x_1^i, ..., x_k^i\}$, with $s = \prod_v s_v$ denoting the number of possible individual states. Let T be the set of all such possible individual states.

In an economic system individual members are *distinguishable*, in the sense that two members with the same individual characteristics $\mathbf{x^i}$ can be understood to exist distinctly from each other and independently of that shared individual state. We may thus cast the economic system as having s^N mathematically conceivable, distinguishable micro-level configurations of the system's N members across each of the s possible coarse-grained individual states. Each micro-level configuration can be represented by a $k \times N$ matrix γ whose column vectors are given by each individual state, $\mathbf{x_n}$.

Naturally, not all conceivable micro-level configurations can occur. The functioning of the system defines a *phase space* Γ consisting of all matrices γ representing the micro-configurations that are phenomenologically possible—i.e., those generated by the functioning of the system. This phase space may be expressed in relation to subspaces Γ_v, each containing all possible micro-level configurations of the N members of the system across the s_v coarse-grained values taken by degree of freedom X_v. Formally, $\Gamma \subseteq \Gamma_1 \times \Gamma_2 ... \times \Gamma_k$.

Coarse graining also allows a distinctive characterization of the system's *macroscopic* state, based on the number of members n_i occupying each of the s individual states. Formally, the system's observed macroscopic state can be described by histogram vectors or frequency distributions $\mathbf{f} = \mathbf{f}_X$,[12] given by $\frac{1}{N}\{n_1, n_2, ..., n_s\}$. The functioning of the system can be understood to define a space or statistical manifold Φ containing all macroscopic states $\mathbf{f}_X$ the system may occupy. The marginal distributions $\mathbf{f}_{X_v}$ over the s_v possible values for an individual degree of freedom X_v are given by $\sum_{x_j \neq x_v} \mathbf{f}_X$. Marginal distributions offer partial descriptions of the system's macroscopic state and can be understood to occupy spaces Φ_v.

[11] A salient discussion of the problems with the latter was provided by Kirman, 1992.

[12] When the domain over which a frequency vector is defined is unambiguously evident, the simple $\mathbf{f}$ will be used. Where there is ambiguity, the subscript will specify the relevant domain.

3.2 Micro-Level Multiplicities and Entropy

Entropy is a conceptual tool allowing formal characterisations of the relationship between micro-level phase spaces Γ and the corresponding sets of accessible macro-level states Φ, defined over any set X of individual degrees of freedom. Every micro-level configuration $\gamma \in \Gamma$ supports a unique macro-state or frequency distribution $\mathbf{f}$. But each macroscopic frequency distribution can be supported by a number of distinct micro-configurations. In fact, the number $\mathbb{W}_{\mathbf{f}}$ of distinct micro-configurations supporting a macro-level state $\mathbf{f}$ under which each bin i contains n_i of the system's N members can be specified. It is given by the multinomial coefficient, which measures the number of ways N distinct elements can be arranged into such a pattern,

$$\mathbb{W}_{\mathbf{f}} = \frac{N!}{\prod_i n_i!} \tag{7}$$

The entropy $H_{\mathbf{f}}$ of a macro-state $\mathbf{f}$ defined over possible values of X, sometimes expressed as $H(X)$, is formally given by the average logarithmic measure of its *multiplicity* $\mathbb{W}_{\mathbf{f}}$ across Γ, which by Stirling's Approximation in the large-N limit can also be expressed as the expected value of $-\log f_i$,

$$H_{\mathbf{f}} = H(X) \equiv \frac{1}{N} \log \mathbb{W}_{\mathbf{f}} = -\sum_x f_x \log f_x \tag{8}$$

Entropy can be understood as a combinatorial heuristic enabling logical inquiry into the functioning of systems of this kind. It can also be understood as a measure of the informational content of a macroscopic state $\mathbf{f}$, given equivalently by either the uncertainty an observer has about the micro-configuration of a system in that state, or by the heterogeneity or diversity in the values of X taken by individual members of the system.

3.2.1 A Combinatorial Heuristic

Entropy is a useful measure for a very simple reason. Higher-entropy macro-states are more common across the distinguishable micro-configurations $\gamma \in \Gamma$ a system may occupy. For systems with $N >> s$, the combinatorial dominance of the distribution $\mathbf{f}^*$ achieving maximum entropy over all other macroscopic states in Φ is overwhelming. So much so that as the system evolves and occupies different micro-configurations over time, it is almost certain that it will remain in *statistical equilibrium* at that macro-state. This is not a phenomenological hypothesis. It is a combinatorial fact, applicable to any system whose micro- and macro-level states may be appropriately described by sets like Γ and Φ, as defined above.

This conclusion can guide the iterative process of observational inquiry into the functioning of such systems. If an observer has a set of knowledge, beliefs, or hypotheses G suggesting that the functioning of a large-N system keeps it within a set Φ^G of macroscopic states, and they have no reason to believe any micro-configuration $\gamma \in \Gamma$ is more likely than others, they should expect to observe macroscopic behavior in line with the state $\mathbf{f}^* = \mathbf{f}^*(G)$ that maximizes entropy over Φ^G. That is because that state is generated by the greatest number of distinct and equiprobable micro-configurations permitted by the system's functioning under G. To expect anything else means either that knowledge different from G is being considered or that expectations are not logically consistent. The distribution $\mathbf{f}^*$ is the distribution consistent with the observer's G that is maximally non-committal toward knowledge that observer does not have. This is the *Principle of Maximum Entropy* (PME).

It is important to note that the PME is not a behavioural hypothesis and is entirely independent of the elements in set G, and of the functioning of the system at hand—or its *semantics*.[13] In fact, if macroscopic behaviour at variance with $\mathbf{f}^*$ is observed, the PME suggests G is either incomplete or wrong, informing subsequent inquiry.[14] What the Principle offers is a distinctive and logically robust way to link certain types of knowledge, beliefs, or hypotheses about the functioning of a system and what basic combinatorial considerations imply about its observable macroscopic states.

3.2.2 A Measure of Information

Entropy is useful well beyond the PME and analysis of systems at statistical equilibria. It offers a measure of the informational content of a macroscopic state $\mathbf{f}$: A quantification, in informational units,[15] of both an observer's uncertainty about the individual states occupied by members of a system, and of the heterogeneity and diversity of those states across the entire population. Consider these in turn.

If $\mathbb{W}_{\mathbf{f}}$ denotes the multiplicity of a macro-state $\mathbf{f}$, we can understand $\xi = \lceil \log \mathbb{W}_{\mathbf{f}} \rceil$ as the smallest number of informational units needed to enumerate or uniquely designate each of the possible micro-configurations in which a system in that state may find itself.[16] We may thus understand the entropy of a system's macroscopic state as an average, informational measure of how much uncertainty we have about the micro-configuration of a system when all we know is that it is in that macro-state. Put differently, entropy provides a measure in informational units of our average uncertainty about the micro-state of an individual member of a system when all we know is that they belong to a population at $\mathbf{f}$.

Entropy also has a related interpretation as an informational measure of heterogeneity in the values of $\mathbf{x}$ taken by members of the system. As can be easily ascertained from 7 and 8, entropy achieves its minimum value of zero for distributions where all individual members are concentrated

[13] As put by Golan, 2018.

[14] For a formal discussion, see Jaynes, 1979, including the derivation of the Entropy Concentration Theorem.

[15] Bits, dits, nats, etc, depending on the base b of the logarithms being used.

[16] That is because ξ information-bearing units each with b distinct states can be used to encode up to b^ξ distinct elements, and $\log_b b^\xi = \xi$.

in a single individual state. All distributions of that kind have a multiplicity of one. It should also be obvious from those two expressions that the multiplicity (and thus the entropy) of macro-states increases as members of a system move from comparatively high to comparatively low occupancy states.[17] The less concentrated or more heterogeneous individual values of $\mathbf{x}$ are across the system, the higher the entropy. At the limit, entropy is maximised when all individuals are uniformly spread across all s possible individual states, where its value reaches $\log s$. As a result, $H_{\mathbf{f}} \in [0, \log s]$.

A number of contributions have also emphasised how entropy can be understood as an index of the diversity in individual states in a population.[18] To see this formally, consider that for every $\mathbf{f} \in \Phi$ defined over s bins there is a unique $\sigma_{\mathbf{f}}$, for which,

$$H_{\mathbf{f}} = \log \mathbb{W}_{\mathbf{f}}^{\frac{1}{N}} = \log \sigma_{\mathbf{f}} \tag{9}$$

It is thus possible to think of a distribution or macro-state $\mathbf{f}$ defined over s bins as informationally equivalent to a uniform distribution defined over $\sigma_{\mathbf{f}} \in [1, s]$ bins, in the sense that both exhibit the same measure of observer uncertainty and heterogeneity. Since the definition in 9 implies that $\sigma_{\mathbf{f}}^N = \mathbb{W}_{\mathbf{f}}$, this equivalence can be expressed differently: The distribution $\mathbf{f}$ has the same multiplicity as the number of conceivable micro-configurations for a system where N members may each take on $\sigma_{\mathbf{f}} \leq s$ distinct individual states. As a result we may understand $\sigma_{\mathbf{f}}$ as a measure of the effective number of individual states available to each member of a system in a macro-state $\mathbf{f}$. Entropy may thus be understood as an index of the diversity of values for X effectively available to members of the system under that distribution.

In what follows, the informational measure of heterogeneity, observer uncertainty, and diversity given by the entropy of a macroscopic state or distribution will be simply referred to as the *information* in that state, distribution, or degrees of freedom. The properties of the information hint at its possible role as the foundation for distinctively useful measures on inequality. It offers a quantification of the dissimilarity in the individual states taken by members of a system. Because it depends only on the frequencies $\mathbf{f}$ with which each individual state is occupied, it is independent of the nature or geometry of the set T of those states. And because it offers a formal measure of observer uncertainty, it can also be related to efforts to characterise the functioning of the system in question. This raises the possibility of developing informational measures of inequality in distributions of any X with a clear relationship to the functioning of the systems generating them.

To understand how entropic functionals can be taken as measures of inequality of that kind, it is necessary to consider carefully multivariate generalisations of entropy and their probabilistic

[17] Formally, if we start from a macro-state or distribution $\mathbf{f^0}$ and move a single individual from state j to state k we arrive at a distribution $\mathbf{f^1}$ whose multiplicity $\mathbb{W}_{\mathbf{f^1}}$ can be characterised relative to the multiplicity of the original distribution: $\mathbb{W}_{\mathbf{f^1}} = \mathbb{W}_{\mathbf{f^0}} \frac{n_j}{n_k+1}$. Clearly, the multiplicity and entropy increase iff $n_j > n_k + 1$, that is, if the movement in question involves an individual moving to a lower-occupancy state than their prior state.

[18] See MacArthur, 1965 for an early application in biology, Patil and Taillie, 1982 for a more general statistical treatment, as well as Jost, 2006.

content.

4 Theil's Limiting Analytical Choice

Before turning to that discussion, a few observations on the limiting analytical choice Henri Theil made when defining his two indices of income inequality are in order. Seen from the perspective of the framework developed above, Theil took "units of income" as the basic members of economic systems, and effectively considered individuals as one-dimensional "states" those units of income may occupy. This poses a number of problems that greatly limit the analytical usefulness of Theil's indices. Those problems follow from the fundamental fungibility of money and the wholly arbitrary and conventional nature of any unit used to account for monetary flows and balances.

Equivalent denominations of money are phenomenologically indistinguishable from one another—their lack of individual economic distinction is in fact inherent to their ability to function as a money. Even in a hypothetical setting where monetary circulation involves the exchange of physical tokens representing a single monetary unit, all such monetary units are perfectly interchangeable: Switching two equivalent physical pieces of money between two individuals while changing nothing else implies no change in the micro-level state of the economic system, its future, or its history.

From the standpoint of the functioning of an economic system, all micro-level configurations of different units of money or income across individuals supporting the same overall distribution or macroscopic state $\mathbf{y}$ are indistinguishable from each other. What is relevant to the evolution of an economic system is how much money or income each individual has, not what exact "pieces" of money or income they have. All micro-level configurations supporting the same $\mathbf{y}$ are in fact the same micro-level state. Mathematically, all macroscopic states $\mathbf{y}$ have a multiplicity of one. The phase space Γ of all possible micro-configurations of units of income or units of money across individuals collapses onto a set equivalent to the set of all possible macro states, Φ.

This collapse does not prevent Theil's indices from providing sound measures of income inequality. But it ensures that they do not offer the correct measure of the multiplicity of micro-configurations supporting a macroscopic state, or of the heterogeneity, uncertainty, and diversity contained in it. That greatly limits their usefulness as tools to characterise income-generating processes.

Using multinomial statistics to describe distributions of indistinguishable entities is generally inappropriate. This was first identified by Satyendra Bose, 1924 in relation to the application of conventional multinomial statistics to analysis of systems made up of particles we now eponymously call Bosons. Bosons occupying the same energy state are indistinguishable. In such cases, the multinomial measure of multiplicity $\mathbb{W}_{\mathbf{f}}$ overestimates the number of distinguishable micro-level configurations supporting a macro-state $\mathbf{f}$, which is correctly measured correctly by Bose-Einstein statistics. The use of multinomial measures to describe distributions of indistinguishable monetary units results in the same kind of error.

Fortunately, the present case does not require the development of new statistics. That is because taking any given denomination of money or income as the unit of economic analysis poses a second and more fundamental problem. As readily evident in contemporary credit-monetary systems, where money consists primarily of entries in bank balance sheets, it is entirely meaningless to speak of discernible individual "units of money" within any given monetary balance. Monetary units of account are entirely arbitrary and conventional. They have no phenomenological significance. Monetary balances change as a result of and in denominations varying according to the interactions between economic agents. In this limited sense, "units of money" are much like units of energy—both are human accounting abstractions used to keep track of changes in the relevant system. Much like energy does not exist independently of physical entities whose state we characterise by an energy level, money balances do not exist independently of the agents holding them. Both units of energy and units of money are patently inappropriate units of analysis.

These difficulties can be readily avoided by taking the approach the previous section outlines. Individual agents are well-defined, inherently distinguishable, and exist independently of their economic state. Their income over any given time period is but one of several individual degrees of freedom defining that state. This simple application of the approach Gibbs' took in his canonical ensembles results in entropy functionals offering correct informational measures of the multiplicity of macroscopic states. That lays the foundation for a non-parametric measure of inequality of opportunity in processes generating any individual economic outcome like income.

5 Mutual Information and Its Probabilistic Content

The formal foundation for the informational index of income inequality is given by the multivariate generalisation of entropy defined by the *co-information* between sets of individual degrees of freedom. Measures of co-information have a clear probabilistic content, defined in terms of a specific symmetry in the processes generating macroscopic distributions or states. In social systems, this symmetry gives formal, mathematical expression to notions of inequality of opportunity. The co-information can also be successively decomposed across sub-groups into additive elements with well-defined content as informational measures of statistical association between individual degrees of freedom. As a result, it offers formal bases for a measure of inequality in the distribution of any given degree of freedom that also captures key aspects of the processes shaping that distribution.

The easiest way to define the co-information and to understand its properties is to consider their simplest and most readily intuitive instance: the *mutual information*, to which this section turns.

5.1 Mutual Information

The mutual information is a concept closely related to entropy that offers very general, nonlinear measures of the phenomenological association between sets of individual degrees of freedom implicit the joint frequency distributions of individuals across their possible values.[19]

5.1.1 Definition

Formally, consider two mutually exclusive sets of individual degrees of freedom in a system: a quantity of primary interest Y and a set of covariates X. The mutual information between Y and X is defined by the difference between the entropy in the marginal distribution of, say, Y and the average entropy of the distributions of Y conditional on values of X,

$$I(Y,X) = H(Y) - H(Y|X) = H(Y) - \sum_{x} f_x \, H(Y|x) \tag{10}$$

The mutual information is the average reduction in heterogeneity in individual values of Y when we move from considering their distribution across all individuals in the system to considering their distributions organised into groupings defined by the values x taken by each individual. It is thus also the reduction in an observer's uncertainty about the system's micro-configuration when they move from observing only the marginal distribution $\mathbf{f}_Y$ to observing the conditional distributions $\mathbf{f}_{Y|x}$—how much we learn, measured in informational units, about individual values of y when we learn about their measure of x. As the name suggests, it is a measure of the information that is shared between two sets of degrees of freedom.[20]

It is trivial to show that the mutual information takes on its minimum value of zero when the conditional distributions $\mathbf{f}_{Y|x}$ have the same entropy as the population-wide, marginal distribution of Y. It increases as the conditional distributions $\mathbf{f}_{Y|x}$ become more organised or less heterogeneous than the population-wide, marginal distribution of Y. And in settings where $H(Y) \leq H(X)$, the mutual information reaches its maximum conceivable value of $H(Y)$ when those conditional distributions have zero entropy. In those cases, all of the heterogeneity or diversity in individual values of Y is associated with heterogeneity or diversity in values of X, ensuring that knowledge of the latter leaves an observer with no uncertainty about the former.

[19] See dos Santos and Wiener, 2019 and dos Santos and Wiener, 2020 for the motivation of the use of different measures of mutual information in observational inquiry into complex economic systems based on the types of data typically available to economists.

[20] Fano, 1961.

5.1.2 An Informational Measure of Association

Simple manipulation underlines that the mutual information is symmetric,[21]

$$I(Y,X) = H(Y) + H(X) - H(Y,X) \tag{11}$$

From this expression it is also possible to show how the mutual information offers a general measure of phenomenological association between sets of degrees of freedom. Using the definitions in 8 and 10 the mutual information can be expressed as,

$$I(Y,X) = \frac{1}{N}\log\frac{\mathbb{W}_{\mathbf{f}_Y}\mathbb{W}_{\mathbf{f}_X}}{\mathbb{W}_{\mathbf{f}_{Y,X}}} \tag{12}$$

This measure captures the extent to which the functioning of a system creates interdependences between individual values for Y and X. If that functioning ensures that the two sets of degrees of freedom are independent, the number of micro-configurations supporting their joint distribution is simply the product $\mathbb{W}_{\mathbf{f}_Y}\mathbb{W}_{\mathbf{f}_X}$ of the number of micro-configurations supporting their respective marginal distributions: For each of the $\mathbb{W}_{\mathbf{f}_Y}$ micro-configurations sustaining $\mathbf{f}_Y$, there would be $\mathbb{W}_{\mathbf{f}_X}$ possible micro-configurations over values of x available to the system. But to the extent that the functioning of the system establishes relationships between measures of y and x, there will be generally fewer than $\mathbb{W}_{\mathbf{f}_X}$ possible micro-configurations over values of x compatible with any of the $\mathbb{W}_{\mathbf{f}_Y}$ possible micro-configurations over values of y. As a result, $\mathbb{W}_{\mathbf{f}_{Y,X}}$ will be smaller than $\mathbb{W}_{\mathbf{f}_Y}\mathbb{W}_{\mathbf{f}_X}$. The mutual information quantifies in informational units this "shortfall" in phase-space volume defined by interdependences in the functioning of the system.

Unlike covariances, the mutual information is defined exclusively by the relative occupancies $\mathbf{f}_{Y,X}$, not the particular locations in T over which they occur. As such, it is independent of the form taken by the interrelationships involved, or from any measure of distance on T or its topology. It can be defined and calculated for any mix of coarse-grained quantitative or categorical degrees of freedom, and will capture the systemic effects of any form of association between values of X and Y within and across individuals in the system.

5.2 Probabilistic Content

The combinatorial foundation of entropic functionals bestows clear probabilistic content upon them.[22] That content ensures that the general, non-parametric measure of association given by the mutual information $I(Y,X)$ is also related to the probability that the processes generating individual values of Y exhibit an important symmetry: That each of the conditional distributions $\mathbf{f}_{Y|x}$ is the result of n_x independent drawings from the population-wide, marginal distribution $\mathbf{f}_Y$.

[21] Using the fact that the definition of entropy ensures that $H(Y|X) + H(X) = H(Y,X)$.

[22] See Niven, 2007.

The mutual information can also be related to a Chi-Squared test-statistic for the hypothesis that this symmetry was present in the processes generating $\mathbf{f}_{Y,X}$

5.2.1 Mutual Information as Probabilistic Measures

Consider taking N independent drawings from a probability distribution $\mathbf{q}$ defined over s states. Let $\pi(\gamma_{\mathbf{f}}, \mathbf{q}; s, N)$ denote the probability that those drawings result in a specific micro-configuration $\gamma_{\mathbf{f}}$ of individuals supporting a macroscopic frequency distribution $\mathbf{f} = \frac{1}{N}\{n_1, ..., n_s\}$. Formally,

$$\pi(\gamma_{\mathbf{f}}, \mathbf{q}; s, N) = \prod_{i=1}^{s} q_i^{n_i} \tag{13}$$

Since there are $\mathbb{W}_{\mathbf{f}}$ such configurations $\gamma_{\mathbf{f}}$, the probability $P(\mathbf{f}, \mathbf{q}; s, N)$ that the drawings will result in a macroscopic frequency distribution $\mathbf{f}$ will be given by,

$$P(\mathbf{f}, \mathbf{q}; s, N) = \pi(\gamma\mathbf{f}, \mathbf{q}; s, N)\ \mathbb{W}_{\mathbf{f}} = \frac{N!}{\prod_i n_i!} \prod_{i=1}^{s} q_i^{n_i} \tag{14}$$

Treading the same path leading to equation 8 leads us to an expression relating the KL divergence $D(\mathbf{f}||\mathbf{q})$ to this probability,

$$\frac{1}{N}\log P(\mathbf{f}, \mathbf{q}; s, N) = H_{\mathbf{f}} + \frac{1}{N}\log \pi(\gamma\mathbf{f}, \mathbf{q}; s, N) = -D(\mathbf{f}||\mathbf{q}) \tag{15}$$

The expression in 15 leads to the widely understood relationship between the entropy of a distribution $\mathbf{f}$ and the probability $P(\mathbf{f}, \mathbf{u}; s, N)$ that that distribution will be the result of N independent drawings from a uniform distribution $\mathbf{u}$ over s states,

$$\log s - H_{\mathbf{f}} = -\frac{1}{N}\log P(\mathbf{f}, \mathbf{u}; s, N) \tag{16}$$

This is the relationship used to establish an identity between Theil's indices and the conjugate pairs of KL divergences discussed above. In line with the discussion in the previous section, note that this relationship is not generally true in systems whose members are indistinguishable.

We may now turn to the probabilistic content of the mutual information. From its definition in 10 and that of the KL divergence in 2, a long-understood identity follows,

$$I(Y, X) = D(\mathbf{f}_{Y,X}||\mathbf{f}_Y\mathbf{f}_X) = -\frac{1}{N}\log P(\mathbf{f}_{Y,X}, \mathbf{f}_Y\mathbf{f}_X; s, N) \tag{17}$$

The mutual information offers a negative, logarithmic measure of the probability that the distribution $\mathbf{f}_{Y,X}$ arose from processes under which X and Y are statistically independent.

Expressing $D(\mathbf{f}_{Y,X}||\mathbf{f}_Y\mathbf{f}_X)$ in 17 in terms of conditional distributions results in two useful identities. The mutual information between degrees of freedom Y and X is in fact the average measure of the informational divergence between each of the conditional distributions $\mathbf{f}_{Y|x}$ and the marginal distribution of Y,

$$I\left(Y,X\right)=\sum_{x}f_{x}\,D\left(\mathbf{f}_{Y|x}||\mathbf{f}_{Y}\right) \tag{18}$$

Application of 15 to each term in this expression and simple manipulation yields what may be a new expression of the probabilistic content of the mutual information,

$$I\left(Y,X\right)=-\frac{1}{N}\log\prod_{x}P\left(\mathbf{f}_{Y|x},\mathbf{f}_{Y};s_{Y},n_{x}\right)\equiv-\frac{1}{N}\log\mathbb{P}\left(\mathbf{f}_{Y|X}||\mathbf{f}_{Y}\right) \tag{19}$$

where $\mathbb{P}\left(\mathbf{f}_{Y|X}||\mathbf{f}_{Y}\right)\equiv\prod_{x}P\left(\mathbf{f}_{Y|x},\mathbf{f}_{Y};s_{Y},n_{x}\right)$ is the probability that the distributions $\mathbf{f}_{Y|x}$ for each and every x sub-group are the independent result of n_x independent drawings from the marginal, population-wide distribution $\mathbf{f}_Y$.[23]

Here the mutual information emerges as a probabilistic measure of systematic differences across sub-groups defined by measures of X in the processes generating values of Y. It is this intuitive property of all measures of informational intersection that allows their use as measures of inequality of opportunity in the determination of individual values of any Y.

5.2.2 Mutual Information as a Test Statistic

Finally, consider the hypothesis that an observed joint distribution $\mathbf{f}_{X,Y}$ is the result of processes where all the conditional distributions $\mathbf{f}_{Y|x}$ are realisations of the same, population-wide marginal distribution $\mathbf{f}_Y$. This is equivalent to the hypothesis that the observed joint distribution is the result of processes under which X and Y are statistically independent.

As shown in Appendix A, if the hypothesis is true, $NI\left(Y,X\right)$ approximately follows the Chi-Squared distribution with $\left(s_Y-1\right)\left(s_X-1\right)-1$ degrees of freedom. That measure may thus be taken as a test statistic for the hypothesis that the processes generating individual values of Y exhibit a simple symmetry across X sub-groups, ensuring that all $\mathbf{f}_{Y|x}$ are realisations of independent drawings from $\mathbf{f}_Y$. In analysis of social systems, where we are often concerned with the presence of systematic differences across sub-groups defined by measures of X in the processes generating values of Y, this defines a test statistic for the presence of equality of opportunity across X sub-groups in the determinations of Y. The test is positive and independent of any normative assessment of X as a basis for differentiation in any economic outcome Y.

[23] The last two terms in argument for $\mathbb{P}$ are omitted for simplicity as they are implied by the first two.

6 Information, Interactions, and Decompositions

The notion of informational overlap, measure of association, and probabilistic content of the mutual information can be generalised to multivariate settings by measures of *co-information* between individual degrees of freedom. That generalisation casts univariate measures of the entropy in the marginal distribution of an individual degree of freedom $H(Y)$ and bivariate measures of mutual information $I(Y,X)$ as first- and second-order instances of a broader measure of informational overlap between any number of individual degrees of freedom, respectively.[24] The co-information also defines formal structures of additive sub-group decompositions of the information in the distribution of Y across and within sub-groups defined by sets of its covariates. Those decompositions generalise the probabilistic relationship in 19 and offer formal, informational characterisations of the associations and interactions involved in the processes shaping Y and its covariates.

6.1 Entropy as a Degeneracy

The entropy in the marginal distribution of any given Y can be usefully understood as a degenerate case of the mutual information. Since $H(Y|Y) = 0$, the *own information* of Y given by $I(Y,Y) \equiv I(Y)$ is simply its entropy. It is a measure of the informational overlap Y has with itself.

This opens a new perspective on the probabilistic content of univariate measures of entropy. The own information defines a degenerate instance of 19: An instance where where each individual n is understood as a sub-group. That is equivalent to considering incomes conditional on each individual's identity, ensuring each "sub-group" or conditional distribution of Y is given by a Kronecker Delta function $\delta(y)$ centred at the individual's state y_n,

$$I(Y,Y) = \frac{1}{N}\sum_n \log P\left(\delta(y_n), \mathbf{f}_Y, N, 1\right) \tag{20}$$

Since the probability of getting a Kronecker Delta function at any given y from a single drawing from $\mathbf{f}_Y$ is f_y, and since there are n_i individuals at each y_i income level, this is clearly equal to the entropy of Y.

The entropy of a marginal distribution $\mathbf{f}_Y$ can be understood as a negative, logarithmic measure of the probability that all of the (degenerate) individual distributions of Y that sustain it are the result of independent drawings that population wide distribution Further manipulation leads to a more precise statement of this,

$$I(Y,Y) = H(Y) = -\sum_y f_y \log f_y = -\frac{1}{N}\sum_y \log f_y^{n_y} = -\frac{1}{N}\log \pi\left(\gamma_{\mathbf{f}_Y}, \mathbf{f}_Y, s, N\right) \tag{21}$$

The entropy of any distribution $\mathbf{f}_Y$ can be related to the probability that the specific micro-

[24] Timme et al., 2014.

level configuration actually supporting that macroscopic state arises as a result of N independent drawings from the distribution $\mathbf{f}_Y$. It offers a probabilistic measure of systematic differences in processes conditioning values of Y across individuals: A negative, logarithmic measure of the probability that those processes offered all individuals the same underlying probabilities $\mathbf{f}_Y$ over all possible values of Y. Unlike the long understood expression of the probabilistic content of entropy in 16, this new characterisation defined without reference to any other arbitrary distribution. It is a property of a distribution itself, defined by a simple symmetry across all individual members of a system.

6.2 Multivariate Generalisations

The co-information is one of several possible multivariate generalisations of the mutual information. Those generalisations are all ultimately defined by the *conditional information*, which is most simply defined for three sets of individual degrees of freedom Y, X, and Z.

6.2.1 The Conditional Information

In that setting, the conditional information measures the informational overlap between Y and X or how much we learn about one of them from observation of the other, when a third set of degrees of freedom Z is already known. Formally,

$$I\left(Y,X|Z\right) = H\left(Y|Z\right) - H\left(Y|\left\{Z,X\right\}\right) = \sum_{z} \mathbf{f}_z \, I\left(Y,X|z\right) \tag{22}$$

The conditional information contains two distinct measures of information.[25] First, there is information shared between Y and X that is not contained in Z. Second, there is information shared between Y and the *combination* of X and Z that is not contained in either X or Z alone. This association between Y and the irreducible combination $\{X,Z\}$ is known as the *synergistic information*. While the formal definition of $I\left(Y,X|Z\right)$ allows calculation of the conditional information from joint distributions $\mathbf{f}_{Y,X,Z}$, there is presently no generally accepted way to identify and measure the synergistic information.[26] Phenomenologically, the conditional information captures the average informational overlap between Y and X within each Z sub-group, as evident in the last term of 22.

As shown in Appendix A, the conditional information $I\left(Y,X|Z\right)$ also defines a test statistic for the presence of a specific symmetry in the processes generating $\mathbf{f}_{Y,X,Z}$: that in those processes the distributions $\mathbf{f}_{Y|x,z}$ are the result of independent drawings from the distribution $\mathbf{f}_{Y|z}$, for all x and z. If this hypothesis is true, $N\ I\left(Y,Z|X\right)$ follows a Chi-Squared distribution with $\left(s_Y-1\right)\left(s_X-1\right)\left(s_Z-1\right)-1$ degrees of freedom.

[25]See Williams and Beer, 2010, for instance.

[26]See Griffith, 2014, for instance.

6.2.2 The Joint Information

The conditional information also defines several useful ways to generalise the measure of informational overlap given by the mutual information for any setting with a degree of freedom of interest Y and any collection $\mathcal{V} = \{\mathcal{V}_1, \mathcal{V}_2, ..., \mathcal{V}_k\}$ of distinct, covariate degrees of freedom considered singly: The *joint information*, the *total correlation*, and the *co-information*.[27]

The joint mutual information captures all of the associations the elements of $\mathcal{V}$ jointly have with Y. In the simplest, trivariate setting where $\mathcal{V}$ contains X and Z it is defined formally as,

$$I\left(Y,\{X,Z\}\right) = I\left(Y,X\right) + I\left(Y,Z|X\right) \tag{23}$$

This is a measure of all the information or heterogeneity shared between Y and all degrees of freedom in $\{X,Z\}$ taken jointly. It includes four distinct components: Information that is shared by Y and X alone, but not by Z; information that is shared by Y and Z alone, but not by X; the synergistic information; and the *redundant* information that can be acquired about Y from observation of *either* X or Y. It should be obvious from application of 19 that the joint information offers a negative, logarithmic measure of the probability that the conditional distributions $\mathbf{f}_{Y|X,Z}$ were the result of independent drawings from $\mathbf{f}_Y$. Along similar lines, Appendix A shows that if those conditional distributions are in fact the result of such independent drawings, $NI\left(Y,\{X,Z\}\right)$ is approximately equal to a variable following the Chi-Squared distribution with $\left(s_Y - 1\right)\left(s_X s_Z - 1\right)$ degrees of freedom.

The general, multivariate form of the joint mutual information between Y and $\mathcal{V}$ is given by,

$$I\left(Y,\{\mathcal{V}\}\right) = I\left(Y,\mathcal{V}_1\right) + I\left(Y,\mathcal{V}_2|\mathcal{V}_1\right) + I\left(Y,\mathcal{V}_3|\mathcal{V}_1,\mathcal{V}_2\right) + ... + I\left(Y,\mathcal{V}_k|\mathcal{V}_1,\mathcal{V}_2,...,\mathcal{V}_{k-1}\right), \tag{24}$$

which captures the intersection between Y and all elements of $\mathcal{V}$ taken jointly—including all synergies and redundancies between elements of the latter in their associations with the former.

6.2.3 The Total Correlation

The joint information defines a further, useful multivariate generalisation of the mutual information: The total correlation. The total correlation is a generalised measure of association among any number of sets of individual degrees of freedom.[28] It captures all interactions involving Y and $\mathcal{V}$, including all redundancies and all synergies between all combinations of degrees of freedom. That includes interactions among the elements of $\mathcal{V}$. This results in a measure that generalises the notion of statistical association in 12, by capturing the shortfall in a system's conceivable phase-space volume due to all interactions between the degrees of freedom in question.

[27] See the extensive, summary discussion in Timme et al., 2014.

[28] Watanabe, 1960.

For three degrees of freedom, the total correlation is given by,

$$TC\left(Y,X,Z\right)=I\left(Y,X\right)+I\left(Z,\{Y,X\}\right)=\frac{1}{N}\log\frac{\mathbb{W}_{\mathbf{f}_Y}\mathbb{W}_{\mathbf{f}_X}\mathbb{W}_{\mathbf{f}_Z}}{\mathbb{W}_{\mathbf{f}_{\{Y,X,Z\}}}} \tag{25}$$

With some manipulation this can be expressed probabilistically,

$$TC\left(Y,X,Z\right)=-\frac{1}{N}\log\left\{\mathbb{P}\left(\mathbf{f}_{Y|X}||\mathbf{f}_Y\right)\mathbb{P}\left(\mathbf{f}_{Y|Z}||\mathbf{f}_Y\right)\mathbb{P}\left(\mathbf{f}_{X|\{Y,Z\}}||\mathbf{f}_{X|Y}\right)\right\} \tag{26}$$

The total correlation offers a negative, logarithmic measure of the probability that three symmetries are independently present in the processes generating values Y, X, and Z: That all X sub-groups faced the same set of possible Y values, that all Z sub-groups also faced the same set of possible Y values, and that within each Y sub-group all Z sub-groups faced the same set of possible X values.

The general expression of the total correlation for any set $\mathcal{V}$ generated alongside Y, is given by,

$$TC\left(Y,\mathcal{V}\right)=I\left(Y,\mathcal{V}_1\right)+I\left(\mathcal{V}_2,\{Y,\mathcal{V}_1\}\right)...+I\left(\mathcal{V}_k,\{Y,\mathcal{V}_1,...\mathcal{V}_{k-1}\}\right). \tag{27}$$

It is trivial to show how this may be expressed in terms of phase-space volumes,

$$TC\left(Y,\mathcal{V}\right)=\log\frac{\prod_i\mathbb{W}_{\mathbf{f}_{\mathcal{V}_i}}}{\mathbb{W}_{\mathbf{f}_{\{\mathcal{V}\}}}}, \tag{28}$$

and how it can be related to a generalised probabilistic expression along the lines of 26.

6.3 The Co-Information

While both the joint information and the total correlation have a wide range of possible uses in work on complex economic systems, neither of them provide satisfactory bases for the decomposition of the information in the marginal distribution of a target degree of freedom Y. The total correlation between Y and any set $\mathcal{V}$ includes informational overlaps among the elements of $\mathcal{V}$ that are unrelated to Y. And the joint information between Y and any set $\mathcal{V}$ is effectively the mutual information $I\left(Y,\{\mathcal{V}\}\right)$ between Y and a single, compound individual degree of freedom composed of the union of those covariates. That limits its usefulness in grappling with the interplay between the elements of $\mathcal{V}$ in the determinations of Y, which is one of the central areas of interest when considering multivariate decompositions of any measure of inequality.

The co-information offers a more useful, multivariate notion of informational overlap between Y and $\mathcal{V}$.[29] It is a measure of informational intersection between degrees of freedom when each of them is considered singly. It's content can be most readily understood in its simplest, trivariate

[29] Due to Matsuda, 2000 and Bell, 2003, the co-information is closely related to the notion of *interaction information* first proposed by McGill, 1954 and developed more recently by Jakulin and Bratko, 2008.

instance $I(Y,X,Z)$,

$$I(Y,X,Z)=I(Y,X)-I(Y,X|Z)=I(Y,X)+I(Y,Z)-I(Y,\{X,Z\}) \tag{29}$$

Here $I(Y,X,Z)$ is defined by how much is learned about Y from X, net of how much is learned about Y from X when Z is already known. It may also be understood as the information shared between Y and X across the entire population, net of the average measure of their shared information within each Z sub-group.

The final term in 29 helps underscore a crucial point in dealing with multivariate informational intersections. $I(Y,X,Z)$ can be thought of as the sum of how much we learn about Y from X and how much we learn about it from Z, net of how much we learn about it from considering $\{X,Z\}$ jointly. It is a *net* measure of informational overlap, given by the difference between the informational redundancy and the informational synergy in X and Z in their association with Y. While this opens a possibility many find counterintuitive—settings where those synergies are greater than those redundancies, ensuring the co-information is negative—its probabilistic content is clear, and may be expressed dually as,

$$I(Y,X,Z)=-\frac{1}{N}\log\frac{\mathbb{P}\left(\mathbf{f}_{Y|X}||\mathbf{f}_{Y}\right)}{\mathbb{P}\left(\mathbf{f}_{Y|X,Z}||\mathbf{f}_{Y|Z}\right)}=-\frac{1}{N}\log\frac{\mathbb{P}\left(\mathbf{f}_{Y|X}||\mathbf{f}_{Y}\right)\mathbb{P}\left(\mathbf{f}_{Y|Z}||\mathbf{f}_{Y}\right)}{\mathbb{P}\left(\mathbf{f}_{Y|X,Z}||\mathbf{f}_{Y}\right)} \tag{30}$$

$I(Y,X,Z)$ is equivalently the probability that X and Y are independent within Z subgroups net of the probability that they are independent across the entire population, and the probability that X and Y are jointly independent from Y net of the probability that they are independently independent from Y.

The general, multivariate form of the co-information $\mathcal{I}(Y,\mathcal{V})$ can be defined recursively, by a succession of subtractions of measures of conditional information along the same lines as 29,

$$\mathcal{I}(Y,\mathcal{V})\equiv I(Y,\mathcal{V}_1,...,\mathcal{V}_{k-1})-I(Y,\mathcal{V}_1,...,\mathcal{V}_{k-1}|\mathcal{V}_k), \tag{31}$$

This generalised measure of informational intersection can be expressed iteratively, in terms of all subsets V of $\mathcal{V}$,[30]

$$\mathcal{I}(Y,\mathcal{V})=\sum_{V\in\mathcal{V}}(-1)^{|V|+1}I(Y,\{V\}) \tag{32}$$

With 12, this becomes an expression involving measures of phenomenological association defined

[30] Appendix C shows the equivalence between this definition and 31.

by phase-space volumes,

$$\mathcal{I}(Y, \mathcal{V}) = \frac{1}{N} \log \prod_{V \in \mathcal{V}} \left(\frac{\mathbb{W}_{\mathbf{f}_Y} \mathbb{W}_{\mathbf{f}_V}}{\mathbb{W}_{\mathbf{f}_{Y,V}}} \right)^{-1^{|V|+1}} \tag{33}$$

It may also be expressed as a generalised statement of the probabilistic result in 30,

$$\mathcal{I}(Y, \mathcal{V}) = \frac{1}{N} \log \prod_{V \in \mathcal{V}} \mathbb{P}\left(\mathbf{f}_{Y|\{V\}} || \mathbf{f}_Y\right)^{-1^{|V|}} \tag{34}$$

involving a series of probabilities that various conditional or sub-group distributions of Y were effectively the result of drawings from the marginal distribution of Y.

6.4 Decompositions of Information

The generalised measure of informational overlap given by the co-information defines a structure of sub-group decompositions of the own information or entropy in the marginal distribution of any degree of freedom Y. The elements in those decompositions have well-defined probabilistic and phenomenological content.

To see this formally, start from the definition of mutual information, which implies,

$$I(Y) = I(Y, X) + I(Y|X) = I(Y, X) + \sum_x f_x \, I(Y|x) \tag{35}$$

The information in the distribution of any Y can be decomposed into an average measure of the information in Y within each x sub-group, and the information in Y shared with the distribution of X across its possible values. Co-information exhibits first-order decomposability across sub-groups.

The same is true for higher-order instances of the co-information, which provide a multivariate generalisation of the notion of across-group shared heterogeneity or information. From its trivariate expression in 29, it follows that,

$$I(Y, X) = I(Y, X, Z) + I(Y, X|Z) = I(Y, X, Z) + \sum_z f_z \, I(Y, X|z) \tag{36}$$

The information shared between Y and X can be decomposed into the average informational association between Y and X within sub-groups defined by values of Z, and into the net informational overlap between X, Y, Z.

This result points to an important consideration in dealing with higher-order decompositions of information across sub-groups that is not widely understood: The need to account for the presence of informational synergies. In the last term of 36, the association between Y and X within Z sub-groups necessarily includes the effects of synergistic interactions between X and Z in relation to Y. Those effects involve Y and the irreducible combination $\{X, Z\}$. That same measure of synergistic information enters as a subtraction in $I(Y, X, Z)$, ensuring that $I(Y, X)$ rightly contains no synergistic information. But in estimates we can obtain from any distribution $\mathbf{f}_{Y,X,Z}$ of both terms in this decomposition it is not possible to isolate the measure of synergistic information present in each of them.[31] Interpretation of the phenomenological content of the decomposition elements in 36 needs to account for the additive and subtractive presence of synergistic information in them.

It is now possible to consider the general decomposition of the own co-information $\mathcal{I}(Y, Y) \equiv \mathcal{I}(Y)$ across a set $\mathcal{V}$ of covariate degrees of freedom,

$$\mathcal{I}(Y) = \mathcal{I}(Y|\mathcal{V}_1) + \mathcal{I}(Y, \mathcal{V}_1|\mathcal{V}_2) + ... + \mathcal{I}(Y, \mathcal{V}_1, ..., \mathcal{V}_{k-1}|\mathcal{V}_k) + \mathcal{I}(Y, \mathcal{V}) \tag{37}$$

The information in any degree of freedom Y can be decomposed into structured series of within-group measures of co-information, plus the generalised, multivariate measure of cross-group heterogeneity in Y given by $\mathcal{I}(Y, \mathcal{V})$. Each element in those decompositions has a clear phenomenological content as an informational measure of association, including all redundancies and synergistic effects. And each element has a clear probabilistic expression, defined by the probabilities that the processes generating values of Y effectively offered a series of sub-groups the same distribution of possible Y values. This offers a robust basis for conceptualising and measuring inequalities of opportunity in the determinations of individual values of any Y in socio-economic systems.

7 The Informational Index of Income Inequality

The foregoing discussion enables the formal definition of a new, informational index of inequality based on measures of co-information, to which this section turns. While the index is motivated in relation to inequality of income, its broader applicability to any set of economic outcomes should be self-evident.

Consideration of the index's definition, as well as its univariate, bivariate, and general multivariate instances, establishes its content as a positive, probabilistic measure of inequality of opportunity in income-generating processes, and as an informational measure of the associations and interactions those processes define. It also establishes the decomposability of its measures into additive structures of within- and across-group measures.

These features enable the index to open innovative lines of conceptual and empirical inquiry into the nature and normative content of patterns of income differentiation in decentralised market economies, some of which are illustrated in the next, concluding section of the paper. They also help identify the probabilistic content of the the index of segregation proposed by Theil and Finezza,

[31] Dissatisfaction with this has motivated the push for different approaches to the multivariate decomposition of information. See Timme et al., 2014; Griffith, 2014 for example.

1971, and define a non-parametric generalisation of the Kitagawa-Oaxaca-Blinder decomposition. After establishing these results, the section turns to the index's mathematical properties as a measure of income inequality.

7.1 An Informational Measure of Inequality

Consider an economic system with a population of N individuals whose economic micro-states are partly described by their income Y over a given time period, and by a set $\mathcal{V}$ of k covariates of income, taken singly. Individual incomes y_n take on values over a finite domain $[0, y_{max}]$, coarse-grained into s_Y income ranges of equal measure. In general, $\mathcal{V}$ contains a combination of categorical and quantitative individual characteristics taking on values over finite domains coarse-grained so as to define a total of $s_{\mathcal{V}}$ possible values v.

Let T denote the set of the $s = s_Y \times s_{\mathcal{V}}$ possible individual states (y, v), and let $\Gamma = T^N$ be the space of possible micro-configurations of the system. The relative occupancy of each of the s individual-state "bins" defines the joint distribution $\mathbf{f}_{Y,\mathcal{V}}$, which may be understood as a description of the system's macroscopic state. Let the set or statistical manifold containing all macroscopic states that can be supported by Γ be denoted by Φ.

7.1.1 Definition

For any macroscopic state $\mathbf{f}_{Y,\mathcal{V}}$, define the *informational index of income inequality* $\mathcal{Q}_Y(V)$ across any subset $V \subseteq \{\mathcal{V} \cup Y\}$ of individual degrees of freedom as,

$$\mathcal{Q}_Y(V) \equiv \mathcal{I}(Y, V) \tag{38}$$

Where $\mathcal{I}(Y, V)$ is the co-information between Y and the elements of V.

Each subset V defines a specific instance of the index. The *order* of any instance V is given by the number of single individual degrees of freedom in $\mathbf{f}_{Y,V}$. Formally, it is equal to $|V \cup Y|$.

7.1.2 A Positive Measure of Inequality of Opportunity

The index $\mathcal{Q}_Y(V)$ is defined exclusively in terms of values for relative occupancies in $\mathbf{f}_{Y,V}$. As such, it offers a purely informational measure of inequality. Its values do not hinge on any notion of location or distance in T; it does not require specification of any notion of social welfare defined over the elements of Γ, and it is defined without reliance on any arbitrary distribution in Φ as a standard of "fairness." This independence ensures the probabilistic and phenomenological content of all of the index's instances is independent of all micro-level details of income-generating processes.

In this the informational index differs from extant measures of income inequality like Theil's indices or the Gini Coefficient. All three of those measures are founded on a normative preference for a specific subset of distributions in Φ: equal distributions of income. Each of those indices gives that preference a distinct semantic expression over T as specific measures of relative income in that set. In contrast, the informational index of income inequality makes no normative presumptions. It can be understood as a positive measure of inequality.

As already suggested by the foregoing discussion, all instances of the index give simple and very general formal expression to the notion of *ex ante* inequality proposed by Friedman, 1962. It offers probabilistic measures of the possibility of differences in the effective "lotteries" faced by each individual and by groups of individuals in income-generating processes. The index offers a natural, informational measure of inequality of opportunity in those processes. A measure that is independent of any normative assessment of what may or may not be morally acceptable as bases of differences in opportunities for individual incomes.

7.2 First-Order, Univariate Instance

The first-order instance of the index is given by the degenerate, univariate case where $V = Y$. It provides an informational measure of inequality in the marginal or population-wide distribution of income $\mathbf{f}_Y = \frac{1}{N}\{n_1, ..., n_{s_Y}\}$.

7.2.1 Formal Specification

Formally, the index's first-order instance $\mathcal{Q}_Y(Y)$ is given by,

$$\mathcal{Q}_Y(Y) = \mathcal{I}(Y) = -\frac{1}{N}\log\pi\left(\gamma_{\mathbf{f}_Y}, \mathbf{f}_Y; s, N\right) = -\frac{1}{N}\sum_y \log f_y^{n_y} \tag{39}$$

As the final two terms in this expression show, $\mathcal{Q}_Y(Y)$ can be related to the probability that the micro-configuration of incomes $\gamma_{\mathbf{f}_Y}$ supporting the distribution $\mathbf{f}_Y$ is the result of independent individual drawings from that distribution. Here the index offers a very natural, probabilistic measure of inequality of opportunity in income-generating processes across all individuals in a population. It may also be understood as a measure of the phenomenological significance of individual identity in those processes.

7.2.2 Inequality in Processes, not Outcomes

A simple example helps illustrate the distinctive measure of inequality captured by $\mathcal{Q}_Y(Y)$, and its contrast with conventional, normative measures of inequality.

Consider two distributions defined over two income levels $\{0, A\}$: a "King-of-the-Castle" distribution, where all but one of a society's N members have zero income, and where the one remaining individual has a level of income $A > 0$; and a "Rotten-Egg" distribution, where only one individual has zero income, with all others receiving A.

Both distributions exhibit the same pattern of relative occupancy, with $\frac{N-1}{N}$ of the population in one income level, and $\frac{1}{N}$ in the other. As a result, they yield the same measure of $\mathcal{Q}_Y(Y) = \log(N) - \frac{N-1}{N}\log(N-1)$. From an informational standpoint, both distributions are entirely equivalent. They are sustained by micro-configurations with the same probability $\pi(\gamma_{\mathbf{f}_Y}, \mathbf{f}_Y; s, N)$.

In contrast, common normative stances result in very different assessments for each distribution. Any welfare metric defined over absolute income levels in T would most likely distinguish between both distributions. The same is true for Theil and Gini indices. In fact, the "King of the Castle" distribution achieves the maximum possible value for all three of these indices, while their measures for the "Rotten Egg" distribution tends to zero for large N.

Behind these stark differences lies the simple fact that inequality of opportunity is not defined over income outcomes *per se*. It is defined by differences in the processes generating those outcomes. In this first-order instance, it captures differences in those processes across all individuals in a population, or differences by individual identity.

7.3 Second-Order Instances and Basic Decompositions

Higher-order instances of the index allow deliberate consideration of the decomposition of total or population-wide inequality of opportunity across different elements or covariates making up individual identities. That in turn enables inquiry based on any normative standard specifying the kinds of individual characteristics that may or may not be acceptable as bases for systematic differentiation of income-generating processes.

The bases for work of that kind can be readily understood in relation to the second-order instances of the index, their probabilistic and phenomenological content, and the decompositions they define.

7.3.1 Definition, Probabilistic, and Phenomenological Content

Second-order instances of the index are defined by V consisting of a single covariate X. Formally,

$$\mathcal{Q}_Y(X) = \mathcal{I}(Y, X) = -\frac{1}{N}\log\mathbb{P}\left(\mathbf{f}_{Y|X}||\mathbf{f}_Y\right) \tag{40}$$

Here the index measures how unlikely it is that the distributions of income for each and every X sub-group arose as the independent result of generative processes offering their members the set $\mathbf{f}_Y$ of possible income levels. It is a probabilistic measure of inequality in the opportunities income-generating processes offered sub-populations defined by X.

$\mathcal{Q}_Y(X)$ also defines a test statistic for the hypothesis that income-generating processes offered equivalent opportunities to all those x sub-populations. If that hypothesis is true, $N\mathcal{Q}_Y(X)$ follows a Chi-Squared distribution with $(s_Y - 1)(s_X - 1) - 1$ degrees of freedom.[32] It can also be understood as a measure of phenomenological association between income and X, given by their informational overlap. This measure is defined without any reference to the topology of the set T of micro-states, measures of distance therein, or to any specification of the processes shaping the associations it measures. This is particularly valuable given the irreducible complexity of income-generating processes.

Instead of requiring suppositions about the details and formal expressions of those processes, $\mathcal{Q}_Y(X)$ is a measure of association based on a simple, parsimonious contention: If the elements of X are not associated with income, we should expect the myriad complex and cumulative processes conditioning individual incomes to result in the same conditional distributions of income across each X sub-group. Conversely, differences among those conditional distributions are the systemic expression of the association between incomes and X. As 18 makes clear, $\mathcal{Q}_Y(X)$ is an average, informational measure of those differences. It may also be taken as a formal measure of how much is learnt about income from observation of individual values of X, as implied by 10 and 12.

It is a very general characterisation of the inequality in the distribution of Y associated with inequality in the distribution of X.

7.3.2 A Conceptual Coincidence and Decompositions

$\mathcal{Q}_Y(X)$ offers a unified mathematical expression for two economic notions of income inequality: Differences in the lotteries of possible income levels facing different individuals or groups of individuals, and the presence of associations between incomes and the individual characteristics defining those groups. Under information-theoretic measures, those two notions coincide and correspond formally to the same aspect of any distribution $\mathbf{f}_{Y,X}$.

This conceptual coincidence is analytically powerful. It bestows clear content to the decomposition of the measure of income inequality provided by $\mathcal{Q}_Y(Y)$. Starting with the simplest case, it follows from the definition of the mutual information that,

$$\mathcal{Q}_Y(Y) = \mathcal{Q}_Y(X) + \sum_x f_x \, \mathcal{Q}_Y(Y|x) \tag{41}$$

The inequality of opportunity in the marginal distribution of income $\mathbf{f}_Y$ can be linearly decomposed into the average piecewise inequality of opportunity within each x sub-group, and the inequality of opportunity across X sub-groups. This decomposition does not suffer from the ambiguity Shorrocks, 1980 identified. Removing all income inequality not associated with X alone amounts to ensuring all the conditional distributions $\mathbf{f}_{Y|x}$ have zero entropy. The only inequality remaining after that

[32] Along analogous lines, second-order, conditional instances of the index $\mathcal{Q}_Y(X|Z) = \sum_z \mathcal{Q}_Y(X|z)$ also define a test statistic for the hypothesis that income-generating processes offered equivalent opportunities to all X sub-populations within each and every subgroup z. If that hypothesis is true, $N\mathcal{Q}_Y(X|Z)$ follows a Chi-Squared distribution with $(s_Y - 1)(s_X - 1)(s_Z - 1) - 1$ degrees of freedom.

removal is $\mathcal{Q}_Y(X)$, which is also the measure of inequality removed if all inequality associated with X alone were eliminated. The informational measure of income inequality "due to" a covariate X is also a measure of the phenomenological association between that covariate and income.

7.3.3 Revisiting the Theil-Finezza Index of Segregation

It is worth noting that the second-order instance of the index is entirely analogous to the index of segregation proposed by Theil and Finezza, 1971. That index was advanced in the first instance as a measure of the extent of segregation by race R in the student populations of schools L in any school district g. Formally, that index I_g obeys,

$$I_g \equiv \sum_l \mathbf{f}_l \sum_l \mathbf{f}_{r|l} \log \frac{\mathbf{f}_{r|l}}{\mathbf{f}_r} = D\left(\mathbf{f}_{R,L} || \mathbf{f}_R \mathbf{f}_L\right) = I\left(L, R\right) = \mathcal{Q}_L\left(R\right) \tag{42}$$

and is equivalent to the mutual information between student race and school allocation.

This equivalence ensures that the index of segregation may also be understood as a measure of inequality of opportunity $\mathcal{Q}_L(R)$–in this case, in the assignment of children of different social groups to schools in a district. The probabilistic result in 40 applies to the index of segregation I_g, which like all measures of mutual information also defines a test statistic for the hypothesis that the processes shaping school placement have not discriminated children by their race. All empirical estimates of I_g may thus be transformed into robust, non-parametric statistical tests for the presence of discrimination.

The equivalence also shows how the Gibbs-Jaynes combinatorial framework defining informational index can grapple with the different applications of information theory Theil and his collaborators were pursuing in the 1960s and 1970s in a conceptually unified manner.

7.4 Multivariate Generalisations

Decomposition of the informational measure of income inequality into linearly additive sub-group elements with equivalent probabilistic and phenomenological content can be usefully generalised for all higher-order instances of the index.

7.4.1 Third-Order Instances and Kitagawa-Oaxaca-Blinder Decompositions

Third-order instances of the index define two useful results. First, $\mathcal{Q}_Y(X, Z)$ is the intersection between the informational measure of income inequality and the informational intersection between the two covariates X and Z. It offers a measure of the inequality of opportunity in income-generating processes associated with the overlap between those covariates, given by the information that is redundant across all of them, net of their synergistic interactions.

That overlap can also be expressed probabilistically,

$$\mathcal{Q}_Y(X,Z) = \frac{1}{N}\left\{\log \mathbb{P}\left(\mathbf{f}_{Y|X,Z}||\mathbf{f}_Y\right) - \log \mathbb{P}\left(\mathbf{f}_{Y|X}||\mathbf{f}_Y\right)\mathbb{P}\left(\mathbf{f}_{Y|Z}||\mathbf{f}_Y\right)\right\} \tag{43}$$

This is a measure of the probability of equality of opportunity in income-generating processes across all (X,Z) sub-populations, net of the probability that those processes independently offered equal opportunities to all X and all Y sub-populations.

Second, $\mathcal{Q}_Y(X,Z)$ defines a decomposition of income inequality across values of covariate X,

$$\mathcal{Q}_Y(X) = \mathcal{Q}_Y(X,Z) + \sum_z f_z\ \mathcal{Q}_Y(X|z) \tag{44}$$

This an informational generalisation of the the Kitagawa-Oaxaca-Blinder decomposition. It formally separates the informational measure of income inequality associated a covariate X into two parts: That which resuls from income inequality by values of X within each Z sub-group, and that which is associated with the intersection between X and Z in income-generating processes.

The informational decomposition in 44 has two advantages over conventional decompositions based on estimation of differences between "average effects" within the terms of linear regression specifications. First, it is non-parametric and valid for any distributional form $\mathbf{f}_{Y,X,Z}$. Second and more significantly, it forces careful consideration of the full range on interactions involved in any decomposition of this kind. The measure of income inequality by X within Z groups inherently includes synergistic interactions between the three covariates. As a result, those effects enter as an addition within the last term in 44. They also enter as a subtraction in the term $\mathcal{Q}_Y(X,Z)$ offering a net measure of income inequality bound up with the intersection of X and Z. Interpretation of within- and across- group elements in decomposition of this kind requires attention to the inherent presence of synergistic effects in income-generating processes.

7.4.2 The General Decomposition

With these specifications it is possible to characterise the general, multivariate decomposition of $\mathcal{Q}_Y(Y)$. For any $k > 1$, the measure of income inequality across $\{\mathcal{V}_1, ..., \mathcal{V}_{k-1}\} = \mathcal{V} \setminus \mathcal{V}_k$ can be expressed as,

$$\mathcal{Q}_Y(\mathcal{V} \setminus \mathcal{V}_k) = \sum_{v_k} \mathbf{f}_{v_k}\ \mathcal{Q}_Y(\mathcal{V} \setminus \mathcal{V}_k|v_k) + \mathcal{Q}_Y(\mathcal{V}) \tag{45}$$

$\mathcal{Q}_Y(\mathcal{V} \setminus \mathcal{V}_k)$ can always be understood as the sum of the average income inequality across $\mathcal{V} \setminus \mathcal{V}_k$ within $\mathcal{V}_k$ sub-groups, and the measure of income inequality across all elements of $\mathcal{V}$.

Iteration of 45 yields a powerful result: A formal decomposition of the informational measure of income inequality in a population into structured sums of within-group measures of that inequality, plus its measure across all covariates defining those sub-groups:

$$\mathcal{Q}_Y\left(Y\right) = \sum_{v_1} \mathbf{f}_{v_1}\,\mathcal{Q}_Y\left(Y|v_1\right) + \sum_{v_2} \mathbf{f}_{v_2}\,\mathcal{Q}_Y\left(\mathcal{V}_1|v_2\right) + ... + \sum_{v_k} \mathbf{f}_{v_k}\,\mathcal{Q}_Y\left(\mathcal{V} \setminus \mathcal{V}_k|v_k\right) + \mathcal{Q}_Y\left(\mathcal{V}\right) \quad (46)$$

The first k terms in this decomposition offer well-defined informational measures of phenomenological association, including all relevant synergistic effects. They also have a clear probabilistic content as measures of inequality of opportunity, which can be obtained formally from application of 34.

7.5 Other Properties

Consider finally the set of properties conventionally associated with measures of income inequality across different instances of the index.

As a univariate measure of entropy in distributions defined over s_Y bins, $\mathcal{Q}_Y\left(Y\right) \in [0, \log s_Y]$. It takes a value of zero only when all individuals have the same income level. The first-order instance of the index is thus non-negative, has an egalitarian zero, and is bounded by a maximum value, occurring when individuals are spread evenly across all possible income levels. As discussed above, third- and higher-order instances of the index may be negative. But they will always be bounded by a maximum value of $\log s_Y$, and take a value of zero when no income inequality is associated with the intersection of covariates in question, taken singly.

Since $\mathcal{Q}_Y\left(V\right)$ is defined exclusively by the relative occupancy of each possible micro-state (y, v), its measure is unchanged under any increase in N that does not change $\mathbf{f}_{Y,V}$. As a result, all instances of the index exhibit population independence. They also also exhibit a generalised version of anonymity: A pairwise switch in income levels between individuals with the same $\mathcal{V}$ does not change the value of any instance of the index.

Scale independence and the Dalton-Pigou Transfer Principle require more careful discussions.

The first-order, population-wide instance of this index is scale independent only in a qualified sense: Its measure is unchanged in the face of any proportional change to all incomes and to the boundaries of all income bins. In more general settings with infinitesimally small or otherwise fixed bin widths, proportional increases of all incomes by an integer factor ϕ result in new distributions of $Z = \phi Y$ with $\mathcal{Q}_Z\left(Z\right) \neq \mathcal{Q}_Y\left(Y\right)$. Under discrete income levels, increases by an integer ϕ effectively spread individuals in any given Y bin across ϕ bins in the domain for Z. That ensures $\mathcal{Q}_Z\left(Z\right) = \log\phi + \mathcal{Q}_Y\left(Y\right)$. Following basic transformation rules leads to the same result for any $\phi > 0$ and distributions defined over a continuous domain.

The fact that $\log\phi$ in wholly independent from the distribution $\mathbf{f}_{Y,\mathcal{V}}$ supports two observations. First, since 45 ensures all higher-order instances of the informational index can be understood in terms of differences between first-order instances, all second- and higher-order instances of the index are in fact scale independent. Second, in applications where scale independence in the measure of univariate inequality is necessary, it is possible to define a version of the index for measures of entropy normalised by the maximum possible entropy the relevant, finite income domain d_Y. This would change only the first-order instance of the index, which becomes, $\bar{\mathcal{Q}}_Y(Y) \equiv \mathcal{Q}_Y(Y) - \log|d_Y|$, which is clearly scale independent.

Finally, $\mathcal{Q}_Y(V)$ does not generally obey the Dalton-Pigou Transfer Principle. However, the index does exhibit a weaker form of the Transfer Principle for unimodal marginal distributions of income. For those distributions, the index offers an informational measure of spread. In such cases, a transfer in income from an individual with a supra-modal level of income to an individual with an infra-modal level of income will never increase $\mathcal{Q}_Y(Y)$. Since observed distributions of income are typically unimodal at reasonable coarse-graining of the income domain, this weaker version of the Principle may be generally valid for a large swathe of empirical distributions of income.

8 New Tools for Normative Work on Income Inequality

The measures of inequality of opportunity provided by the informational index are purely informational, not normative. But by offering coincident, informational quantifications of inequality of opportunity in income-generating processes and of phenomenological associations involved in them, it offers new, robust bases for normative work drawing on ethical notions of what may or may not be acceptable bases for income differentiation. Among several possible applications, this includes innovative approaches to statistical measurement of the effects on incomes of social realities of gender, race, ethnicity, or class background make to measures of income inequality.

8.1 Grappling with Inequality of Opportunity

A number of salient contributions have carefully considered the normative content of various bases for economic differentiation across individuals in decentralised market economies. That includes differentiation by income, which is closely associated with an individual's ability to command social resources in those economies. Most notably, Rawls, 1971 offered a negative thesis concerning the distribution of the burdens and fruits of social cooperation as part of his comprehensive, Liberal scheme of justice as fairness: That *morally arbitrary* characteristics that are not the result of an individual's actions—like parental background, innate physical or mental capacities, sex, race, etc—should *not* be the bases for differentiation in economic outcomes. This notion was developed by "luck egalitarians" like Roemer, 1998,[33] who advanced notions of justice in economic distributions based on the view that *circumstantial characteristics* that are not defined by the agency of conscious individuals should not affect economic outcomes.

Across any such schematisation, the informational index offers a useful conceptual and empirical metric tool. Conceptually, it offers a formal, consistent way to think about inequalities of

[33] See also Dworkin, 2000, Anderson, 1999, Cohen, 1989, and Arneson, 1989, for instance.

opportunity and about associations and interactions defined during income-generating processes. Empirically, the index offers a set of powerful, informational tools to measure the associations between income inequality and any set of covariates understood as normatively objectionable bases for economic differentiation. Those measures and the statistical tests they define are true independently of the details of the relationships involved, and how much researchers know about their mathematical forms. In this the informational index of income inequality (and analogous indices for inequality in the distribution of other economic outcomes) makes a distinctive and useful addition to the extant set of tools used in empirical studies of inequalities of opportunity.[34]

Use of the index as a normative tool demands and in fact helps provide a clear understanding of the limitations of equality of opportunity as a normative maxim, and its situation within broader notions of social justice. Those limitations are evident in at least three broad areas. First and as argued by several contributions,[35] the absence of normatively unjustifiable inequalities in *ex-ante* treatment or opportunities is insufficient to guarantee most commonly accepted notions of fairness in economic outcomes. Certain outcomes in T may be entirely unjustifiable regardless of the processes behind them. Sufficiency in normative assessments of the fairness requires the incorporation of basic quality-of-life standards and associated threshold measures for certain economic outcomes.

By giving a formal measure to the notion of inequality of opportunity, the informational index underlines a second, broader insufficiency of equality of opportunity as a normative maxim. Even in states where all individual outcomes in T are normatively acceptable, there can be important and widely recognised normative differences between distributions with similar measures of inequality of opportunity. This was evident in the index's measures of inequality for the "King-of-the-Castle" and "Rotten Egg" distributions above.

Third and more generally, notions of equality of opportunity *taken by themselves* effectively accept the moral legitimacy of economic differentiation based on the purported coincidence between measures of individual effort and agency on one hand, and individual rewards on the other. This is a very specific, individualist normative contention that is rather common in contemporary "meritocratic" discourses. Several salient thinkers favoured very different notions of justice, including both Liberal and Marxist egalitarians who advanced normative standards fundamentally grounded on a recognition of the pervasive social interdependences between individuals in a decentralised, capitalist economy. Rawls, 1971 considered that over and above equality of opportunity, differentiation in individual social and economic outcomes could only be justified on the basis of its contribution to the collective, social well-being of the community as a whole. Drawing on very different ethical and methodological tenets, Marxists like Lenin, 1999 understood individualised notions of fairness as "bourgeois," and justifiable only inasmuch as scarcity prevents a community from reaching more meaningfully equitable outcomes at any given stage of its historical and economic development.

[34]See Roemer and Trannoy, 2013, for instance.

[35]Hufe et al., 2021; Kanbur and Wagstaff, 2016; Fleurbaey, 1995

For Marxists like him, it is only the historical development of productive capacities that can lay the foundation for a more meaningful standard of social justice, defined explicitly by Marx, 1970: That social production and distribution will be just when material abundance permits all to contribute in proportion to their abilities, and appropriate in proportion to their needs.

These caveats underscore an important point in applications of the informational index developed above. It is most usefully deployed as part of a negative approach to measures of justice in economic outcomes, in settings where we may never come to know the details of the complex processes shaping them. It gives researchers very general measures of systematic differences in those outcomes across groups defined by morally ambiguous or circumstantial individual characteristics. Even as researchers don't know the details involved in income-generating processes, and make contending claims about what constitutes an exhaustive set of principles defining distributive justice, there may be readier agreement on elements of what constitutes distributive *injustice*. The index enables innovative work along those lines. This is illustrated in general, conceptual terms below, and with a numerical example in Appendix C.

8.2 A Conceptual Example

For any given system of normative postulates and principles, it is possible to think about the complex, unknown, cumulative, and largely unobservable processes shaping individual incomes Y in terms of three distinct kinds of factors and their interactions. First, they involve a set of individual characteristics $\mathbb{C}$ that are not acceptable as a basis for economic differentiation. Second, they involve a set of characteristics $\mathbb{E}$ that are in themselves accepted as bases for economic differentiation. Finally, income is also determined by contingencies and other processes and factors that escape even conceptual identification.

In this setting, the total measure of income inequality $\mathcal{Q}_Y(Y)$ that can be informationally accounted for is given by $\mathcal{Q}_Y(\{\mathbb{C}, \mathbb{E}\}) \leq \mathcal{Q}_Y(Y)$, with the difference between those two measures corresponding to irreducible indeterminacy in the distribution of income that escapes characterisation. That total measure of tractable inequality has four components: income inequality associated with $\mathbb{C}$ alone and not associated with $\mathbb{E}$; income inequality associated with $\mathbb{E}$ alone and not associated with $\mathbb{C}$; and two terms defined by the interplay between the two sets—the inequality associated with the informational overalap or redundancies between $\mathbb{C}$ and $\mathbb{E}$ considered singly; and the income inequality due to synergistic interactions between those two sets in income-generating processes.

Conceptually, normative assessments of the first two components are fairly straightforward, by construction. Normative assessments of redundancies and synergies between the two sets of covariates and income requires more careful consideration. What is the normative content of the informational redundancy between $\mathbb{C}$ and $\mathbb{E}$ in their respective associations with inequality of income? Perhaps more interestingly, what are we to make of the synergistic interactions between elements of each set in the determination of a highly consequential economic outcome like income?

Each of these questions also poses practical challenges. What are the analytical limitations and implications for policy discourses of tackling redundancies between $\mathbb{C}$ and $\mathbb{E}$ in the determination of income by "controlling for" elements of $\mathbb{E}$ within the terms of parametric linear regression models, and deeming only "residual" measures of the influence of elements of $\mathbb{C}$ as normatively objectionable? And what are the analytical limitations of seeking to grapple with the synergies between elements of $\mathbb{C}$ and $\mathbb{E}$ in the determination of incomes by using *ad hoc* multiplicative "interaction terms" in linear regression models?

With all of these questions, the informational index of income inequality and the conceptual framework within which it is defined are distinctively helpful.

Practical inquiry always involves observation of subsets $C \in \mathbb{C}$ and $E \in \mathbb{E}$, and observer ignorance about their complements in $\mathbb{C}$ and $\mathbb{E}$. It also often involves C whose elements are morally ambiguous or circumstantial individual characteristics that, (a) are determined prior to processes conditioning income and its covariates, and (b) assign individuals to social groups accorded systematically different treatment during those processes. For example, C may contain social-identity characteristics like gender, race, ethnicity, or class background; while E contains economic characteristics like educational attainment and work experience.

In this type of setting, differences across C sub-groups both in their distributions across values of E, and in the ways the elements of E shape incomes can be understood as expressions of the social systems that define and assign economic significance to the elements of social identity in C. Formally, the redundancy between C and E in the processes conditioning Y can be taken as an expression of the social realities defining gender, race, ethnicity, and class: Those realities fundamentally shape how individuals in different social-identity groups are distributed across different e sub-groups of economic characteristics, and how those differences in turn shape incomes.[36] Along similar lines, the synergistic effects between C and E in income-generating processes can also be understood as part of the social assignment of economic significance to the elements of C. As a result, it is possible to consider that income inequality associated with both the redundancies and synergies between C and E are expressions of the broad *influence* realities of social identity exert over incomes.

Here the informational index of income inequality defines a collection of generally applicable conditions that can inform applied normative inquiry. That work is often concerned with the independent association between elements of E and individual incomes. Grappling with that association requires looking into the conditional information,

$$\mathcal{Q}_Y(E|C) = \sum_c \mathbf{f}_c \, \mathcal{Q}_Y(E|c) \tag{47}$$

This is a measure of how improbable it is that within each C subgroup, members of each E sub-

[36] Along the same lines followed in Roemer, 1998,dos Santos and Wiener, 2020, and Hufe et al., 2021.

group faced the same opportunities for levels of income as the entire C sub-group. It is also a measure of the informativeness of E within C subgroups.

But elements of E may have synergistic associations with elements of C in income generating processes. While those synergies are included in the measure of $\mathcal{Q}_Y(E|C)$ in 47, they should be understood as expressions of the social systems conditioning the effects of E on incomes across C groups. Further, the associations between elements of E and incomes may also be redundant once unobserved characteristics in $\mathbb{C} \setminus C$ are considered. That too would ensure that parts of the overlap captured by $\mathcal{Q}_Y(E|C)$ are in fact attributable to social systems shaping unobserved circumstantial characteristics. The presence of redundancies and synergies ensure that in fact $\mathcal{Q}_Y(E|C)$ is in fact an *upper bound* on the independent influence of any observable set of elements of E exerted over income-generating processes. Put differently, *at least* $\mathcal{Q}_Y(Y) - \mathcal{Q}_Y(E|C)$ informational units of inequality are not attributable to observed economic characteristics E providing normatively acceptable bases for income differentiation.

Conversely, the value of $\mathcal{Q}_Y(C)$ offers a measure of how improbable it is that each and every C sub-group faced the same set of possible income outcomes. And it helps define test statistics for the hypothesis that income-generating processes offered equal opportunities to all C groups, against the hypothesis that they did not.

It is also a measure of the influence the elements of C *alone* exert over incomes. This can be understood as the sum of the direct influence elements of C have over income independently of elements of E, and the redundant information about incomes that is contained in both C and E. Since the latter cannot be measured, it is also useful to think about this influence is as a sum of two quantities: The influence elements of C exert over incomes within each E group, and the inequality of opportunity in incomes across E and C groups, considered singly. Formally and respectively,

$$\mathcal{Q}_Y(C) = \sum_e \mathbf{f}_e \, \mathcal{Q}_Y(C|e) + \mathcal{Q}_Y(C, E) \tag{48}$$

Note that $\mathcal{Q}_Y(C)$ does not include synergistic effects that elements of C may have with elements of E. Those enter into the first term on the right-hand side of 48 as an addition, but enter as a subtraction on the second term. As a result, $\mathcal{Q}_Y(C)$ it is in fact a *lower bound* on the normatively objectionable influence circumstantial characteristics C exert over incomes.

This approach helps fundamentally shift debates concerning statistical measurement of the effects circumstantial characteristics and the social systems that define them have on incomes. Those debates have been largely defined by statistical estimates of parameters in strongly specified models of the formal relationships in T between income and elements of E and C. In such exercises, all estimates and statistical tests are contingent on the empirical purchase of the particular specifications that defined them. What researchers don't know about the details of complex income-generating processes and what they cannot observe gets in the way of developing robust measures of the effects of the elements of C over incomes.

In contrast, $\mathcal{Q}_Y(C)$ is well defined and robust for any income-generating processes, any distributional forms generated by those processes, and any subset E we actually observe. It provides a measures of inequality of opportunity; test statistics for the presence of equality of opportunity; and measures of the phenomenological association between income and the elements of C that are well-defined and readily estimated for any well-populated $\mathbf{f}_{Y,C,E}$. The resulting generality helps change the relationship between knowledge researchers do not have and what they can conclude from what they observe in this kind of work. What is not known no longer undermines their ability to obtain sound estimates of the lower bound for normatively undesirable measures of inequality associated with the observed elements of C. It simply pertains to the unknown details of the mechanisms and associations shaping the inequality of opportunity by elements of C they can measure.

This is but one promising way where the informational index of income inequality may help advance normative work on the nature of income differentiation. It is hoped that alongside the conceptual discussion in earlier sections, these applications spark interest in the index and its application to available data.

References

Anderson, E (1999). "What is the Point of Equality." *Ethics* 109, pp. 287–337.

Arneson, R (1989). "Equality and Equal Opportunity." *Philosophical Studies* 56, pp. 77–93.

Bell, Anthony J. (2003). "The Co-Information Lattice." In: *4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003).* 4. ICA, pp. 921–926.

Blinder, A (1973). "Wage Discrimination: Reduced Form and Structural Estimates." *Journal of Human Resources* 8.4, pp. 436–455.

Bose, S (1924). "Planck's Law and the Hypothesis of Light Quanta." *Zeitchrift fur Phyik* 26.178-181.

Caticha, Ariel and Amos Golan (2014). "An entropic framewokr for modeling economies." *Physica A* 408, pp. 149–163.

Cohen, G (1989). "On the Currency of Egalitarian Justice." *Ethics* 99.4, pp. 906–944.

Conceicao, Pedro, James Galbraith, and Peter Bradford (2000). "The Theil Index in Sequences of Nested and Hierarchic Grouping Structures." *UTIP Working Paper* 15.

Cowell, Frank A. (2006). "Theil, Inequality Indices and Decomposition." *Dynamics on Inequality and Poverty Research on Economic Inequality* 13, pp. 345–360.

Csiszar, Imre (1998). "The Method of Types." *IEEE Transactions on Information Theory* 44.6, pp. 2505–2523.

dos Santos, Paulo and Noe Wiener (2020). "By the Content of Their Character? Comparative Group Income Heterogeneity and Socio-Economic Discrimination." *Journal of Mathematical Sociology* 44.1.

dos Santos, Paulo L. (2020). "Statistical Equilibria in Economic Systems: Socio-Combinatorial or Individualist-Reductionist Characterizations?" *European Physical Journal Special Topics* 229, pp. 1603–1622.

dos Santos, Paulo L. and Noe Wiener (2019). "Indices of Informational Association and Analysis of Complex Socio-Economic Systems." *Entropy* 21.367.

Dworkin, R (2000). *Sovereign Virtue: The Theory and Practice of Equality*. Harvard University Press.

Einstein, Albert (1925). "Quantum Theory of the Monoatomic Ideal Gas." *Sitzungsberichte der Preussischen Akademie der Wissenchaften*.

Fano, R. M. (1961). *Transmission of Information: A statistical theory of communication*. MIT Press, p. 57059.

Farjoun, F and M Machover (1983). *Laws of Chaos: A Probabilistic Approach to Political Economy*. Verso.

Fleurbaey, M (1995). "Equal Opportunity or Equal Social Outcome?" *Economics and Philosophy* 11.1, pp. 25–55.

Foley, Duncan K. (1994). "A Statistical Equilibrium Theory of Markets." *Journal of Economic Theory* 62.2, pp. 321–345.

Friedman, Milton (1962). *Capitalism and Freedom*. University of Chicago Press.

Georgescu-Roegen, N (1971). *The Entropy Law and the Economic Process*. Harvard University Press.

Gibbs, J. Willard (1902). *Elementary Principles in Statistical Mechanics*. Charles Scribner's Sons.

Golan, Amos (2018). *Foundations of Info-Metrics*. Oxford University Press.

Griffith, Virgil (2014). "Quantifying Synergistic Information." PhD thesis. California Institute of Technology.

Hufe, Paul, Ravi Kanbur, and Andreas Pichl (2021). "Measuring Unfair Inequality: Reconciling Equality of Opportunity and Freedom from Poverty." *Review of Economic Studies* forthcoming.

Jakulin, A and I Bratko (2008). "Qunatifying and visualizing attribute interactions." *arXiv:cs/0308002v3*.

Jaynes, Edwin T. (1979). "Concentration of Distributions at Entropy Maxima." In: *E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics*. Ed. by R. D. Rosenkrantz. D. Reidel, Dordrecht.

— (1989). "Delaware Lecture." In: *E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics*. Ed. by R. D. Rosenkrantz. Springer, pp. 87–113.

— (2003). *Probability Theory–The Logic of Science*. Cambridge University Press.

Jost, Lou (2006). "Entropy and Diversity." *Oikos* 113.2, pp. 363–375.

Kanbur, Ravi and Andy Snell (2019). "Inequality Indices as Tests of Fairness." *The Economic Journal* forthcoming.

Kanbur, Ravi and A Wagstaff (2016). "How Useful is Inequality of Opportunity as a Policy Construct?" In: *Inequality and Growth: Patterns and Policy*. Ed. by K Basu and J Stiglitz. Palgrave MacMillan.

Keynes, John Maynard (1921). *A Treatsie on Probability*. Macmillan and Co., Ltd.

Kirman, Alan P. (1992). "Whom or What Does the Representative Individual Represent." *Journal of Economic Perspectives* 6.2, pp. 117–136.

Kitagawa, Evelyn (1955). "Components of a Difference Between Two Rates." *Journal of the American Statistical Association* 50.272, pp. 1168–1194.

Lenin, Vladimir (1999). "The State and Revolution." In: *Collected Works, Volume 25*. Vol. 381-492. Progress Publishers.

MacArthur, R (1965). "Patterns of species diversity." *Biological Review* 40, pp. 510–533.

Mantegna, Rosario and H. Eugene Stanley (2000). *An Introduction to Econophysics and Finance*. Cambridge University Press.

Marx, Karl (1970). "Critique of the Gotha Program." In: *Marx/Engels Selected Works, Volume Three*. Vol. 13-30. Progress Publishers.

Matsuda, Hiroyuki (2000). "Physical nature of higher-order mutual information: Intrinsic correlations and frustration." *Physical Review E* 62.3, pp. 3096–3102.

McGill, W. J. (1954). "Multivariate Information Transmission." *IEEE Trans. Information Theory* 4.4, pp. 93–111.

Niven, Robert (2007). "Combinatorial Information Theory: I. Philosophical Basis of Cross-Entropy and Entropy Combinatorial Information Theory: I. Philosophical Basis of Cross-Entropy and Entropy." Unpublished Manuscript, University of New South Wales.

Oaxaca, R (1973). "Male-Female Wage Differentials in Urban Labor Markets." *International Economic Review* 14.3, pp. 693–709.

Patil, G and C Taillie (1982). "Diversity as a concept and its measurement." *Journal of the American Statistical Association* 77, pp. 548–561.

Pearson, Karl (1900). "X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling." *Philosophical Magazine Series* 5.50:302, pp. 157–175.

Rawls, J (1971). *A Theory of Justice*. Harvard University Press.

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

Roemer, John and Alain Trannoy (2013). "Equality of Opportunity." *Cowles Foundation Discussion Paper* 1921.October.

Rosser, Barkley (2008). "Econophysics." In: *The New Palgrave Dictionary of Economics, 2nd Edition*. Palgrave MacMillan, pp. 1625–1643.

Maximum Entropy Economics (2020) 229.

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

Shannon, Claude E. (1948). "A Mathematical Theory of Communication." *Bell System Technical Journal* 27, pp. 379–423, 623–656.

Shorrocks, A (1980). "The Class of Additively Decomposable Inequality Measures." *Econometrica* 48.3, pp. 613–625.

Stutzer, Michael J (1983). "Toward a Statistical Macrodynamics: An Alternative Means of Incorporating Micro Foundations." *Federal Reserve Bank of Minneapolis Research Department Working Paper* PACS File 3300.242.

Theil, Henri (1971). *Economics and Information Theory*. Elsevier.

Theil, Henri and Anthony Finezza (1971). "A Note on the Measurement of Racial Integration of Schools by Means of Informational Concepts." *Journal of Mathematical Sociology* 1, pp. 187–194.

Timme, Nicholas, Wesley Alford, and Benjamin Flecker (2014). "Synergy, redundancy, and multivariate information measures: an experimentalist's perspective." *Journal of Computational Neuroscience* 36, pp. 119–140.

Watanabe, Satosi (1960). "Information Theoretical Analysis of Multivariate Correlation." *IBM Journal of Research and Development* 4.1, pp. 66–82.

Williams, P.L. and R.D. Beer (2010). "Nonnegative Decomposition of Multivariate Information." *arXiv* 1004.2515.

Yakovenko, Victor (2007). "Econophysics, Statistical Mechanics Approach to." In: *Encyclopedia of Complexity and System Science*. Ed. by RA Meyers. Springer.

9 Appendix A- Derivation of Test Statistics

We are interested in the hypothesis that an observed joint distribution $\mathbf{f}_{X,Y}$ is the result of processes where the conditional distributions $\mathbf{f}_{Y|x}$ are realisations of the same marginal distribution of income $\mathbf{f}_Y$. Call this Hypothesis A.

To consider this formally, start from the basic probabilistic content of mutual-information measures in 19. For natural logarithms, that implies that,

$$\mathbb{P}\left(\mathbf{f}_{Y|X}||\mathbf{f}_Y\right) = \exp - NI\left(Y, X\right) \tag{49}$$

If we express the mutual information as a divergence along the lines of 17, define $\epsilon = \mathbf{f}_X\mathbf{f}_Y - \mathbf{f}_{X,Y}$, and use a second-degree Taylor approximation around $\epsilon = \mathbf{0}$, this becomes,

$$\mathbb{P}\left(\mathbf{f}_{Y|X}||\mathbf{f}_Y\right) \approx \exp - \frac{1}{2}N\sum_{x,y}\frac{\epsilon_{y,x}^2}{f_{y,x}} \tag{50}$$

Along analogous lines to those followed by Pearson, 1900 and Jaynes, 1979, it is possible to transform coordinates in the statistical manifold containing all possible $\mathbf{f}_{Y,X}$ by letting $\chi^2 \equiv N\sum_{x,y}\frac{\epsilon_{y,x}^2}{f_{y,x}}$. This allows an approximate, continuous parametrisation of the statistical manifold as a set of concentric, $((s_Y - 1)(s_X - 1) - 1)$-dimensional spherical shells with radii χ^2, centred at a distribution with $\epsilon = \mathbf{0}$. From this it is possible to characterise, within the terms of the Taylor approximation and up to an integration coefficient, the volume $V_{Y,X}$ of all possible outcomes of N drawings when X and Y are independent falling inside a sphere with $\chi^2 \approx NI\left(Y, X\right)$,

$$V_{Y,X} \sim \int_0^{NI(Y,X)} \chi^{(s_Y-1)(s_X-1)-1} \exp - \frac{1}{2}\chi^2 \; d\chi \tag{51}$$

From this it should be evident that if Hypothesis A is true, $NI\left(Y, X\right)$ approximately follows the Chi-Squared distribution with $(s_Y - 1)(s_X - 1) - 1$ degrees of freedom.

This result can be extended to the conditional information $I\left(Y, X|Z\right) = \sum_z I\left(Y, X|z\right)$, which defines two additional (sets of) hypotheses. Consider first the hypothesis that an observed distribution $\mathbf{f}_{X,Y|z}$ is the result of processes where the distributions $\mathbf{f}_{Y|x,z}$ are realisations of the same distribution of income $\mathbf{f}_{Y|z}$. Call this contention, defined piecewise for a given z, Hypothesis B.

It should be immediately obvious by direct application of the line of reasoning above that if Hypothesis B is true, $\chi_z^2 \equiv n_z\sum_{x,y}\frac{\epsilon_{y,x|z}^2}{f_{y,x|z}} \sim n_zI\left(Y, X|z\right)$ approximately follows Chi-Squared distribution with $(s_Y - 1)(s_X - 1) - 1$ degrees of freedom.

Consider now Hypothesis C—that hypothesis B is true independently for all z. Here we start from the piecewise identity,

$$\mathbb{P}\left(\mathbf{f}_{Y|X,z}||\mathbf{f}_{Y|z}\right) = \exp - n_zI\left(Y, X|z\right) \tag{52}$$

From which it follows that,

$$\mathbb{P}\left(\mathbf{f}_{Y|X,Z}||\mathbf{f}_{Y|Z}\right) = \exp - \sum_{z} n_z I\left(Y, X|z\right) = \exp - NI\left(Y, X|Z\right) \tag{53}$$

A simple transformation allows an approximately continuous parametrisation of the statistical manifold of all $\mathbf{f}_{Y,X,Z}$ as a series of concentric, $(s_Y - 1)(s_X - 1)(s_Z - 1) - 1$-dimensional spherical shells with radii $\mathcal{X}^2 \equiv \sum_z \chi_z^2$, centred at a distribution with $\epsilon = \mathbf{0}$. Along the same lines as above, we conclude that if Hypothesis C is true, the test statistic $NI(Y, X|Z)$ follows a Chi-Squared distribution with $(s_Y - 1)(s_X - 1)(s_Z - 1) - 1$ degrees of freedom.

10 Appendix B - Iterative and Recursive Definitions of Co-Information

This section establishes the equivalence between the iterative definition of co-information in 32 and the accepted, recursive definition in 31. Starting from the former,

$$\mathcal{I}(Y, \mathcal{V}) = \sum_{V \in \mathcal{V}} (-1)^{|V|+1} I(Y, \{V\}) \tag{54}$$

It is evident that when $\mathcal{V}$ consists of a single covariate $\mathcal{V}_1$ this yields the first-order instance of the co-information, $\mathcal{I}(Y, \mathcal{V}_1)$, and that the addition of a second covariate $\mathcal{V}_2$ to $\mathcal{V}$ yields the second-order instance, $\mathcal{I}(Y, \mathcal{V}_1) + \mathcal{I}(Y, \mathcal{V}_2) - \mathcal{I}(Y, \{\mathcal{V}_1, \mathcal{V}_2\})$.

Consider now the co-information $\mathcal{I}(Y, \mathcal{V}, X)$ between any $\mathcal{V}$, Y, and an additional covariate X,

$$\mathcal{I}(Y, \mathcal{V}, X) = \sum_{V \in \mathcal{V} \cup X} (-1)^{|V|+1} I(Y, \{V\}) \tag{55}$$

This sum may be separated into terms corresponding to subsets V involving only elements of $\mathcal{V}$, and terms corresponding to subsets $\bar{V}$ that contain the new covariate X. Letting $\bar{\mathcal{V}}$ denote the set of all $\bar{V}$, 55 becomes,

$$\mathcal{I}(Y, \mathcal{V}, X) = \sum_{V \in \mathcal{V}} (-1)^{|V|+1} I(Y, \{V\}) + \sum_{\bar{V} \in \bar{\mathcal{V}}} (-1)^{|\bar{V}|+1} I\left(Y, \{\bar{V}\}\right) \tag{56}$$

There is a bijection between $\mathcal{V}$ and $\bar{\mathcal{V}} \setminus X$, defined by $\bar{V} = \{V, X\}$. This allows expression of the last term in 56 as a sum over all V plus a single term for $\bar{V} = X$. Noting that $|\bar{V}| = |\{V, X\}| - 1$, this expression is given by,

$$\mathcal{I}(Y, \mathcal{V}, X) = \sum_{V \in \mathcal{V}} (-1)^{|V|+1} I(Y, \{V\}) - \sum_{V \in \mathcal{V}} (-1)^{|V|+1} I(Y, \{V, X\}) + I(Y, X) \tag{57}$$

Using the definition of the joint information to expand the second summand, this becomes,

$$\begin{aligned}\mathcal{I}\left(Y, \mathcal{V}, X\right) = \sum_{V \in \mathcal{V}} \left(-1\right)^{|V|+1} I\left(Y, \{V\}\right) - \sum_{V \in \mathcal{V}} \left(-1\right)^{|V|+1} I\left(Y, \{V\} | X\right) + \\ I\left(Y, X\right) - \sum_{V \in \mathcal{V}} \left(-1\right)^{|V|+1} I\left(Y, X\right)\end{aligned} \tag{58}$$

For any $\mathcal{V}$ the final sum in 58 always yields a single $-I\left(Y, X\right)$, by the properties of the binomial coefficient. That ensures the final two terms disappear, yielding the recursive definition, $\mathcal{I}\left(Y, \mathcal{V}, X\right) = \mathcal{I}\left(Y, \mathcal{V}\right) - \mathcal{I}\left(Y, \mathcal{V}|X\right)$, proving the equivalence by induction.

11 Online Appendix - C Numerical Example

A "synthetic" example can help illustrate how the informational index of income inequality may be applied in work based on work based on observed joint distributions of income and covariates, and the kinds of inferences it enables.

Consider a total population of one million income earners with income levels between zero and $10,000$ monetary units. Suppose further that observation of those incomes is only possible at a coarse graining with bin widths equal to the monetary unit. As a result, there are $10,000$ bins of possible income levels, and the maximum possible measure of entropy is given by $\log 10,000$, or approximately 9.21 if natural logarithms are used.

Suppose this population may be decomposed into four groups, defined by values of two observable characteristics—educational attainment E; and social-identity G, defined for all individuals independently and prior to their income-generating processes. Education may take on a "high" or "low" level, denoted by $e = \bar{e}$ and $\underline{e}$, respectively. And there are two social-identity groups: a historically dominant group with $g = \bar{g}$, and a historically subordinate group with $g = \underline{g}$. Let high-education, dominant-status individuals with $(\bar{e}, \bar{g})$ make up 20 percent of the population; high-education, subordinate-status individuals with $(\bar{e}, \underline{g})$ make up ten percent of the population; low-education but dominant-status individuals with $(\underline{e}, \bar{g})$are 40 percent of the population; and low-education, subordinate-status individuals with $(\underline{e}, \underline{g})$ make up 30 percent.

Suppose further that observed income distributions for these four subgroups are each well described by a Lognormal distribution $\mathcal{L}(y; \mu, \sigma)$, defined for income values y between zero and $10,000$. All four distributions have a shape parameter $\sigma = 1$, but differ in their scale parameters μ. Let those take values of $4, 3, 2,$ and 1 respectively.

Nine groups can be defined in this setting this population. Table 1 shows the distributional form for incomes in each of them, along with its corresponding measure of entropy.

Table 1: Population Groupings, Income Distributions, and their Entropy

Group	Weight	Income Distribution	Entropy
All	1	$0.2\,\mathcal{L}(y;4,1)+0.1\,\mathcal{L}(y;3,1)+0.4\,\mathcal{L}(y;2,1)+0.3\,\mathcal{L}(y;1,1)$	3.995
$\bar{g}$	0.6	$0.33\,\mathcal{L}(y;4,1)+0.66\,\mathcal{L}(y;2,1)$	4.357
$\underline{g}$	0.4	$0.25\,\mathcal{L}(y;3,1)+0.75\,\mathcal{L}(y;1,1)$	3.185
$\bar{e}$	0.3	$0.66\,\mathcal{L}(y;4,1)+0.33\,\mathcal{L}(y;3,1)$	5.144
$\underline{e}$	0.7	$0.57\,\mathcal{L}(y;2,1)+0.43\,\mathcal{L}(y;1,1)$	3.100
$(\bar{g},\bar{e})$	0.2	$\mathcal{L}(y;4,1)$	5.419
$(\underline{g},\bar{e})$	0.1	$\mathcal{L}(y;3,1)$	4.419
$(\bar{g},\underline{e})$	0.4	$\mathcal{L}(y;2,1)$	3.419
$(\underline{g},\underline{e})$	0.3	$\mathcal{L}(y;1,1)$	2.419
Maxent			9.210

From these measures of entropy for group income distributions it is possible to calculate the values for all instances of the informational index of income inequality for this population. Those are listed in Table 2.[37]

Table 2: Index Values

Instance	Definition	Value
$\mathcal{Q}_Y(Y)$	$H(Y)$	3.995
$\mathcal{Q}_Y(G)$	$\mathcal{Q}_Y(Y)-\sum_g f_g\,\mathcal{Q}_Y(Y\|g)$	0.106
$\mathcal{Q}_Y(E)$	$\mathcal{Q}_Y(Y)-\sum_e f_e\,\mathcal{Q}_Y(Y\|e)$	0.282
$\mathcal{Q}_Y(G\|E)$	$\sum_e f_e\,\mathcal{Q}_Y(G\|e)$	0.094
$\mathcal{Q}_Y(E\|G)$	$\sum_g f_g\,\mathcal{Q}_Y(E\|g)$	0.269
$\mathcal{Q}_Y(\{G,E\})$	$\mathcal{Q}_Y(G)+\mathcal{Q}_Y(E\|G)$	0.376
$\mathcal{Q}_Y(G,E)$	$\mathcal{Q}_Y(G)-\mathcal{Q}_Y(G\|E)$	0.012

The total informational measure of income inequality in this population of one million is 3.995. This means the probability that income-generating processes offering each individual the same set of possible income levels, as given by the marginal or population wide distribution of income, result in the observed configurations of individual incomes is $e^{-3.995\times10^6}$ or $10^{-1.73\times10^7}$.

[37] And organised by their effective degree. Note that the conditional and joint instances are effectively second-degree instances of the index.

This total can be decomposed into the average income inequality within each of the two G subgroups, $\mathcal{Q}_Y(Y|G) = 3.889$, and income inequality across G groups, $\mathcal{Q}_Y(G) = 0.106$. As discussed in section 7, the latter measure is a lower bound on total income inequality associated with the social realities defining the groups g.

That implies that the probability of obtaining the distributions of income observed for groups $barg$ and $\underline{g}$ from processes independently offering members of each group the population-wide distribution of possible income levels is at most $e^{-1.06\times10^5} \sim 10^{-46,035}$. If income-generating processes did indeed offer both groups the same set of possible income values, $N\mathcal{Q}_Y(G)$ follows a Chi-Squared distribution with $9,999$ degrees of freedom. The value of that test statistic for the assumed income distributions is 1.06×10^6. The probability that income-generating processes that are equitable across G groups generate the observed distribution is an unfathomably small $4.47 \times 10^{-15,724}$. The evidence of inequality of opportunity across G groups in the generation of incomes is overwhelming. It is also independent of the micro-kinetic details of income-generating processes, the distributional forms they define, the influence of unobserved individual characteristics on income-generating processes, and of our knowledge about all of these things as observers.

Following the discussion in section 7, the maximum measure of inequality that may be associated with education in this case is given by $\mathcal{Q}_Y(E|G) = 0.269$. The total amount of income inequality across the entire population that may be jointly accounted for by values of E and G is $\mathcal{Q}_Y(\{G,E\}) = 0.376$, or about 9.4 percent of total, population-wide inequality. The 0.106 accounted for by G amounts to about 28 percent of the total population inequality that can be accounted for by what is observed in this case. The remaining measure of inequality of opportunity can be attributed to heterogeneity of incomes within (E,G) subgroups.

It is also possible to obtain third-order decompositions of income inequality in this case. This helps cast some light onto the processes involved in defining the significance of G in income-generating processes. The total measure of inequality associated with social identity $\mathcal{Q}_Y(G)$ can be decomposed into two distinct parts: The measure of income inequality associated with g within each education group, $\mathcal{Q}_Y(G|E) = 0.094$; and the measure of income inequality associated with the intersection of education and social identity considered singly, $\mathcal{Q}_Y(G,E) = 0.012$. This is an informational generalisation of the decomposition put forward by Kitagawa, Blinder, and Oaxaca based on estimated linear-regression parameters.

This decomposition in turn defines a third-degree decomposition of the total, population-wide measure of inequality. That total may be though of as the sum of the measure of income inequality within social-identity groups, $\mathcal{Q}_Y(Y|G) = 3.889$; the inequality across social-identity groups within education groups, $\mathcal{Q}_Y(G|E) = 0.094$; and inequality shared by income, education, and social identity, $\mathcal{Q}_Y(G,E) = 0.012$, when each is considered singly.

Finally, it bears noting that these are population-wide measures. As such, they are conditioned by the relative weights of different subgroups in the population as a whole. This underestimates the effects social-identity and education have on groups accounting for relatively small fractions of the population. Take for instance the 0.4 of the total population in the subordinate social group $\underline{g}$. The measure of income inequality associated with that specific social identity, $\mathcal{Q}_Y\left(Y|\underline{g}\right)$ is 0.810. Observing that an individual is a member of the socially subordinate group in this example removes more than 20 percent of our uncertainty about their income. Along analogous lines, the measure of income inequality associated with the social identity and level of education of low-education, subordinate-status individuals is $\mathcal{Q}_Y\left(Y|\underline{g},\underline{e}\right) = 1.57$. Learning somebody is a member of the socially subordinate group with a low level of education eliminates more than 39 percent of our uncertainty about their income.