Article

Untangling Synergistic Effects of Intersecting Social Identities with Partial Information Decomposition

Thomas F. Varley 1,2,* and Patrick Kaminski 1,3

1 School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN 47405, USA
2 Department of Psychology & Brain Sciences, Indiana University, Bloomington, IN 47405, USA
3 Department of Sociology, Indiana University, Bloomington, IN 47405, USA
* Correspondence: tvarley@iu.edu

Abstract: The theory of intersectionality proposes that an individual’s experience of society has aspects that are irreducible to the sum of one’s various identities considered individually, but are “greater than the sum of their parts”. In recent years, this framework has become a frequent topic of discussion both in social sciences and among popular movements for social justice. In this work, we show that the effects of intersectional identities can be statistically observed in empirical data using information theory, particularly the partial information decomposition framework. We show that, when considering the predictive relationship between various identity categories such as race and sex, on outcomes such as income, health and wellness, robust statistical synergies appear. These synergies show that there are joint-effects of identities on outcomes that are irreducible to any identity considered individually and only appear when specific categories are considered together (for example, there is a large, synergistic effect of race and sex considered jointly on income irreducible to either race or sex). Furthermore, these synergies are robust over time, remaining largely constant year-to-year. We then show using synthetic data that the most widely used method of assessing intersectionalities in data (linear regression with multiplicative interaction coefficients) fails to disambiguate between truly synergistic, greater-than-the-sum-of-their-parts interactions, and redundant interactions. We explore the significance of these two distinct types of interactions in the context of making inferences about intersectional relationships in data and the importance of being able to reliably differentiate the two. Finally, we conclude that information theory, as a model-free framework sensitive to nonlinearities and synergies in data, is a natural method by which to explore the space of higher-order social dynamics.

Keywords: partial information decomposition; synergy; sociology; intersectionality; higher-order interactions

1. Introduction

“Intersectionality” refers to the idea that an individual’s experience of social privilege or oppression is a function of how all the various identities that a given person can hold (such as race, sex, class, ability, etc.) intersect, and that the result of these intersecting identities is not directly decomposable into the sum of all identities considered individually. Intersectionality as a framework was first articulated by Kimberle Crenshaw [1,2], who discussed how Black women in particular face distinct forms of marginalization in the context of the American justice system. She argued that pre-existing theoretical frameworks of both feminist and anti-racist scholarship erased the specific experiences of Black women, instead operating under a framework where “All the Women Are White; All the Blacks Are Men” [1]. The particular intersection of Blackness and race has been called misogynoir by Moya Bailey [3] to articulate its status as a unique, irreducible social experience.

From its initial focus on Black feminism and the unique experience of Black women, the field of Intersectionality Studies has expanded [4] and now many different axes of
identity are routinely studied, including sexuality [5,6], class [7], disability [8,9], and/or immigration status and citizenship [10–13]. Outside of academia, intersectionality has gained prominence in popular culture, particularly around issues related to the Black Lives Matter movement, trans and gender-nonconforming rights, and the issue of immigration from South and Central America into the United States. As a consequence of the growing public discussion and activism, there has been increasing interest in how intersectional frameworks could be deliberately incorporated into public policy decisions [14–16]. As theories of intersectionality leave the Academy and begin to influence more mainstream decision-making, it is imperative that the field develops analytic tools that allow for the discussion of intersecting identities in the context of empirical data. Understanding the causal and material impacts of policy given intersectional identities requires the development of robust empirical methodologies that can be used to inform and identify interventions.

Previous researchers have highlighted the difficulties in addressing the issue of synergistic relationships between identities in empirical data: for example Bowleg concisely summarizes the issue as "Black + Lesbian + Woman ≠ Black Lesbian Woman", and discusses the problems with assuming additive relationships, although Boweleg is primarily concerned with analysing qualitative ethnographic data and does not engage with the mathematical issues around super-additive relationships in numerical or categorical data [17]. There has been discussion of how intersectionality can be accounted for by quantitative methods, for example Scott and Siltanen [18] discuss multi-level linear regressions informed by context, and Rouhani provides a gold-standard primer detailing how to distinguish between additive and multiplicative effects using linear regressions [19]. Generally, the standard practice is identifying intersectional effects with the multiplicative interaction term in linear regression and comparing it to the main effect using estimators of prediction error like Akaike’s Information Criteria.

While powerful, these methods have a number of fundamental limitations that can complicate the analysis of complex data. The reliance on parametric models, goodness-of-fit tests, and arbitrary criteria such as the a-level of statistical significance can limit the kinds of relationships the analysis is sensitive to. For example, the reviewed literature makes overwhelming use of linear (Gaussian) assumptions in assessing main and interaction effects, and does not account for non-linear relationships between interacting variables. Another significant concern is that, while linear regressions can compare main and interaction effects, and in doing so account for interactions between identities, it fails to capture the synergistic core of intersectionality: no distinction is made between different kinds of interactions, such as redundant vs. synergistic dependencies. For example, a Black woman plausibly experiences: generic anti-Black racism (independent of sex), generic misogyny (independent of race), and intersectional misogynoir specific to her identities as a Black woman. Furthermore, there may also be "redundant” effects: a cost shared by Black people and women by virtue of being a minority of any type and not specific to either identity. Linear regression models fail to decompose this constellation of effects, as they have no way to rigorously account for, or even acknowledge, differences between redundant and synergistic relationships. This inability has implications for our ability to model the social dynamics that give rise to systemic, higher-order inequalities, as different kinds of generative dynamics may give rise to redundant or synergistic dependencies. For example, redundant dependencies are often reflective of a common-driver effects [20] which are often conditioned out, while synergies represent irreducible, higher-order interactions that cannot be simplified and possess their own structure reflecting multivariate dependencies in the data. To address these issues, we turn to an alternative statistical framework for analysing data: information theory.

Information theory represents an appealing statistical framework with which to tackle the question of intersectionality in quantitative data: in contrast to standard linear regression models, information theory usually does not require making any assumptions about the distribution of the data (normal, Poisson, etc.), making it sensitive to non-linear relationships in data [21]. It is well-known that Gaussian models can fail when the underlying
generative dynamic is strongly non-normal, such as in heavy-tailed distributions (for example, the distribution of wealth in the United States) [22,23], temporally extended human dynamics, such as epidemic spreading [24] and these nonlinearities can lead to erroneous conclusions and incorrect models [25]. Finally, for researchers who do wish to leverage the power of linear models, or work with continuous data, closed-form Gaussian estimators of all major information-theoretic relationships exist: for example, the Pearson correlation is a function of the mutual information between Gaussian variables [21], the partial correlation maps to the conditional mutual information [26], and the Granger Causality is a special case of the more general transfer entropy [27]. There also exist non-parametric, continuous estimators based on K-nearest neighbor relationships [28] for non-linear analysis of real-valued (continuous) data (also see [29], Supplementary Material).

Crucially, information theory is well-equipped to handle the problem of decomposing multivariate relationships in data into synergistic (intersectional) and redundant components using a framework known as partial information decomposition (PID) [30,31] (see Section 2.2 for details), and has been applied in a variety of fields, including interpretable machine learning [32], medical imaging [33], biological neural networks [34,35], ecology [36], evolution [37], as well as to philosophical questions such as the nature of “emergence” [38–40] and consciousness [41]. This interdisciplinary group of results suggests that synergistic relationships “greater than the sum of their parts” are ubiquitous in both natural and human-made systems, so it is natural to hypothesize that they may also exist in social systems.

In this paper, we propose a conceptual link between synergistic, “greater than the sum of their parts” information with intersectionality in sociological data. By applying the partial information decomposition analysis to US census data, we can test for higher-order interactions between multiple identities (race, sex, etc.) on outcomes like income or health status. We hypothesized that statistical synergies indicating the presence of intersectional inequalities (i.e., costs associated with being both Black and female or advantages associated with being both White and male) should be observable in population-level demographic and life-outcome data using information theoretic analysis, and that information theory will out-perform linear regression in discriminating between redundant and synergistic effects. Furthermore, we go on to compare the information-theoretic analysis with standard, multivariate linear models to show that the linear approaches are not equipped to differentiate between different kinds of multivariate dependence.

2. Methods

2.1. Basic Information Theory

Information theory is a mathematical framework that describes how different interacting entities inform on and constrain each-other’s behaviour [21]. Originally developed in the context of theories of communications [42], information theory has become an indispensable tool for the analysis of complex systems [20,43].

2.1.1. Information Theory & Inference

To build an intuition about how information theory can help us understand interactions between different identities, it is helpful to consider simple examples. Imagine selecting a person at random and, knowing nothing about their sex, race, or any other demographic considerations, trying to predict their income. In the absence of any contextual information, the maximally likely income is the expected value of the whole income distribution (in this case $43,308.13 USD, see Section 2.4), although there is going to be variability within the population which makes us uncertain about the accuracy of our estimate (the entropy, see below). Now, suppose that we learn the sex of our randomly selected person (male or female). Knowing this new piece of information will reduce our uncertainty about their likely income by ruling out everyone inconsistent with the target sex. In a fundamental way, this is what it means to say that sex discloses information about income. We can walk through a similar example with the category race. Suppose
we pick a new person at random, and learn their race (White or Black). Similarly, our initial uncertainty about income is reduced, because race and income are correlated. Finally, suppose we learn both the race and the sex of our randomly selected person: knowing the joint state of both predictors will reduce our uncertainty about income the most. If we learn more information from knowing both demographic variables together than we learn from the sum of the two predictors considered independently, then there is some information in the whole that is greater than the sum of its parts—what we call the intersectional synergy. This is a distinct concept from the multiplicative interaction term more commonly used in regression analysis as it does not necessarily assume a particular kind of interaction (multiplicative), but instead is based on the information that is only accessible when considering both identities jointly. As we will see, the information-theoretic perspective allows us to explore other dependencies, such as the difference between learning race and sex compared to learning race or sex (which correspond to synergy and redundancy, respectively). These represent two different kinds of interaction that can be disentangled, while the regression only gives us one kind of interaction.

2.1.2. Entropy & Mutual Information

The core object of study in information theory is the entropy, which quantifies how uncertain we, as observers, are about the state of a variable we are observing. Inference, then, is typically understood as the process of minimizing entropy by understanding how information about the variable in question is disclosed by both its own statistics, and related variables. Given a variable $X$ which can take on different states $x$ drawn from a support set $X$ according to the probability distribution $P_X(x)$, the entropy of the variable is given by:

$$H(X) = - \sum_{x \in X} P_X(x) \log(P_X(x)). \quad (1)$$

If we are observing more than one variable, the joint entropy is an easy generalization: for a pair of variables $\{X, Y\}$, the joint entropy is given by:

$$H(X, Y) = - \sum_{x \in X} \sum_{y \in Y} P_{X,Y}(x,y) \log(P_{X,Y}(x,y)) \quad (2)$$

Which quantities the total uncertainty about the state of both variables simultaneously. From the joint and marginal entropies we can calculate the conditional entropy:

$$H(X|Y) = H(X, Y) - H(Y). \quad (3)$$

This conditional entropy tells us how much uncertainty about the state of $X$ remains after we have “accounted for” the state of $Y$. It is important to note that $H(X|Y) \leq H(X)$: information about the state of $Y$ can only ever reduce our uncertainty about the state of $X$ or, if $X$ and $Y$ are independent, provide no insight.

So far, all the measures that we have discussed have been measures of uncertainty, not measures of information. To quantify information itself, we introduce the mutual information:

$$I(X; Y) = H(X) - H(X|Y). \quad (4)$$

For a visual intuitive aid, see Figure 1.
Figure 1. Mutual information as the intersection of marginal entropies. Mutual information can be understood as the intersection the entropies of two correlated variables. This plot highlights the intuition that \( I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) \).

It is worth unpacking this to build intuition: we begin with our uncertainty about the state of \( Y \), given by \( H(Y) \). From this, we subtract our remaining uncertainty about \( Y \) after the state of \( X \) has been taken into account (\( H(Y|X) \)). The difference of these two quantities is our uncertainty about \( Y \) that is resolved by knowledge of \( X \): the amount of information \( X \) provides about \( Y \). It is important to note that the mutual information is symmetric: \( I(X; Y) = I(Y; X) \). As with the joint entropy, it is possible to calculate the mutual information between a set of variables and a single target. Consider the case of trying to predict the state of a variable \( Y \) based on two predictor variables \( X_1 \) and \( X_2 \). We can calculate \( I(X_1, X_2; Y) \) the same way as above, only treating the joint states of \( X_1 \) and \( X_2 \) as a single macro-variable.

It is crucial that the joint mutual information is not always equivalent to the sum of the individual marginal mutual informations:

\[
I(X_1, X_2; Y) \neq I(X_1; Y) + I(X_2; Y). 
\] (5)

The joint mutual information can be either greater, or less than, the sum of the marginal mutual informations, depending on how correlated information is distributed across the two predictor variables. If \( I(X_1, X_2; Y) < I(X_1; Y) + I(X_2; Y) \) then the total is less than the sum of their parts and consequently \( X_1 \) and \( X_2 \) must share some redundant information. Consider Venn diagram in Figure 2: if the area of two overlapping circles is less than the sum of the areas of both circles considered independently, then there is redundant area shared between them. Conversely if \( I(X_1, X_2; Y) > I(X_1, Y) + I(X_2, Y) \), then the whole is greater than the sum of the parts: there is information about \( Y \) in the joint state of \( X_1, X_2 \) that cannot be extracted from either variable considered independently. In the case of two predictors and a single target variable, the difference between the whole and the sum of the parts (sometimes called the "Whole Minus Sum" integration [44]), is equivalent to a number of other information-theoretic measures, such as the co-information (a trivariate generalization of mutual information) and related to a recently proposed measure of redundancy/synergy bias: the O-information [45]. In the general case, it has also been introduced under the name Redundancy/Synergy Index [46].

Finally, it is entirely possible that a pair of predictor variables contains both redundant and synergistic information: while the difference between the joint and the sum of the marginal mutual informations is a decent heuristic to test if there is any synergy at all in the triad, it does not tell us how much redundant or synergistic information is present, nor what proportion of the total information is accounted for by redundant or synergistic information. For this we need more advanced mathematical machinery.
Partial Information Decomposition

Given a set of predictor variables influencing a single predicted variable, the partial information decomposition framework [30,31] provides the tools necessary to decompose the joint mutual information into redundant, synergistic, and unique “types”. The resulting decomposition is given by:

\[
I(X_1, X_2; Y) = \text{Red}(X_1, X_2; Y) + \text{Unq}(X_1; Y/X_2) + \text{Unq}(X_2; Y/X_1) + \text{Syn}(X_1, X_2; Y). \tag{6}
\]

where \(\text{Red}(X_1, X_2; Y)\) quantifies the information about \(Y\) that can be resolved by observing \(X_1\) OR \(X_2\), \(\text{Unq}(X_1; Y/X_2)\) quantifies the amount of information about \(Y\) is uniquely disclosed by \(X_1\) (in the context of \(X_2\)) and vice versa. Finally, \(\text{Syn}(X_1, X_2; Y)\) quantifies the information disclosed only by the joint state of \(X_1\) AND \(X_2\), and no simpler combination of variables. We can decompose the marginal mutual informations in the same fashion:

\[
I(X_1; Y) = \text{Red}(X_1; Y) + \text{Unq}(X_1; Y/X_2), \tag{7}
\]

\[
I(X_2; Y) = \text{Red}(X_2; Y) + \text{Unq}(X_2; Y/X_1). \tag{8}
\]

The result is a system of three equations, three known quantities (the joint and marginal mutual informations), and four unknown variables (\(\text{Red}, \text{Unq}_1, \text{Unq}_2, \text{Syn}\), referred to as the “atomic” components of the mutual information). If it is possible to identify any one partial information term, then the remainder are trivial. The synergy term can also be written as the difference between the joint mutual information and all the simpler combination of sources:

\[
\text{Syn}(X_1, X_2; Y) = I(X_1, X_2; Y) - \text{Red}(X_1, X_2; Y) - \text{Unq}(X_1; Y/X_2) - \text{Unq}(X_2; Y/X_1). \tag{9}
\]

In this perspective, we can see that synergy is the information about the target that is accessible when observing the whole (both sources) and none of the parts considered independently. In Figure 2, it is equivalent to the violet area of the oval. This definition of synergy as the difference between the joint state of both predictors (AND) and the
redundant information (OR) provides a useful framework with which the PID can be connected to other relevant analyses based on logical conjunctions and disjunctions (see Section 3.1).

2.3. Choosing a Partial-Information Function

Unfortunately, classical Shannon information theory does not provide a function that calculates any of these, and neither does the PID framework itself. Much work has gone into developing such functions, and as yet, there is no consensus within the field of information theory as to the “gold standard”. To ensure the robustness of the concept, we analysed the data using two different functions to ensure that the distributions of redundancy and synergy remained consistent despite different technical starting points. The first was the original measure proposed by Williams and Beer [30] called $I_{\text{min}}$, which defines the redundant information shared by a set of sources about a target as:

$$ I_{\text{min}}(X_1, X_2, \ldots, X_n; Y) = \sum_{y \in Y} P_Y(y) \min_i I(X_i; Y = y). \quad (10) $$

This measure (also called the “specific information”) gives the minimum amount of information that the set of sources discloses about the target. To ensure that our results were robust to the particular free parameters inherent in the PID framework, we replicated our results using the measure of $I_{\text{BROJA}}$ proposed by Bertschinger et al. [47,48]. The $I_{\text{BROJA}}$ method starts with the unique information, rather than the redundant information; defining it as the minimum conditional mutual information possible while holding the marginals constant. To extract the unique information $I_{\text{BROJA}}(X_1; Y/X_2)$, Bertschiner et al., begins by defining the set of all joint distributions of $X_1$, $X_2$, and $Y$ such that the marginals are equivalent to the empirical distribution $P$:

$$ \Delta = \{ Q | P_{X,Y}(x_i, y) = Q_{X,Y}(x_i, y) \forall i \} \quad (11) $$

The unique $I_{\text{BROJA}}$ for a given source is defined as:

$$ I_{\text{BROJA}}(X_1; Y/X_2) = \min_{Q \in \Delta} I_Q(X_1; Y|X_2). \quad (12) $$

Where $I_Q(\cdot; \cdot|\cdot)$ indicates that the mutual information being calculated with respect to the probability distribution $Q(X_1, X_2, Y)$. Once the two unique informations have been calculated, the rest of the terms can be solved with basic algebra. Both $I_{\text{min}}$ and $I_{\text{BROJA}}$ provide strictly positive values for all partial information atoms, allowing for a complete decomposition of the joint mutual information terms. As mentioned above, controversies and discussions remain about the “best” approach to defining partial information functions: both $I_{\text{min}}$ and $I_{\text{BROJA}}$ have their own distinct limitations and particular behaviours. $I_{\text{min}}$ has been criticised for behaving in “counter-intuitive” ways in some edge cases [49], although $I_{\text{min}}$ remains by far the most widely-used redundancy function in analysis (see [32,34,39,50] for a non-exhaustive list). $I_{\text{BROJA}}$ was developed in part in response to the limitations of $I_{\text{min}}$, although it is restricted to the triadic case of two predictors and a single target and cannot be readily generalized to more complex structures, unlike $I_{\text{min}}$. In the main body of the paper we report the results of the $I_{\text{min}}$ analysis, and the results of $I_{\text{BROJA}}$ (where applicable) are available in the Supplementary Information.

2.4. Data & Pre-Processing

For this project, we used a decade of data (2011–2020) from the Annual Social and Economic (ASEC) Supplement (available here: https://www.census.gov/data/datasets/time-series/demo/cps/cps-asec.2020.html, accessed on 9 September 2021), which provides individual-level micro-data on demographic and economic indicators in the United States. We excluded individuals younger than 25 years of age, or older than 65 years of age as they were less likely to be working full-time (children, students, or retired), individuals not
native-born US Citizens, and we only analyzed those who self-identified as “White only” or “Black only” to limit the confounds associated with multi-racial identities. Finally, we only considered individuals who were working in full-time employment since we were looking explicitly at the effects of identities like race or sex on income and the possibility of different levels of employment would represent a possible confound.

To limit the size of the joint probability space, we coarse-grained the income distribution: individuals making $0.00–$27,499 were placed into a Low Income group, individuals making $27,500–$52,499 were placed into a Lower-Middle Income group, individuals making $52,500–$77,499 were placed into Upper Middle Income group, and finally those making more than $77,500 were placed into an High Income group, although we found that our results were generally robust to the exact choice of cutoffs or the number of categories up to a point at which point the joint probability distribution was too sparsely sampled. When considering health outcomes, we did not coarse grain the data, which was reported on a 5-item Likert scale.

2.5. Analysis Pipeline

Performing the PID analysis requires constructing the entire joint-probability distribution for all variables of interest. For example, when untangling the synergistic effects of sex and race on income, it is necessary to compute:

\[ P(Race = \text{Black} \land Sex = \text{Male} \land Income = \text{Low}), \]
\[ P(Race = \text{White} \land Sex = \text{Male} \land Income = \text{Low}), \]
\[ P(Race = \text{Black} \land Sex = \text{Female} \land Income = \text{Low}), \]
\[ P(Race = \text{White} \land Sex = \text{Female} \land Income = \text{Low}), \]
\[ \ldots \]
\[ P(Race = \text{White} \land Sex = \text{Female} \land Income = \text{High}), \]

where \( \land \) is the logical-AND operator.

To accurately estimate these probabilities, it is crucial to ensure that the distribution of joint states of the predictor variables (in this case, race and sex) are evenly distributed: there should be an equal number of Black women, White women, Black men, and White men. If, for example, White men are over-represented and have higher-than-average incomes, which is the case in the full dataset, then that will incorrectly skew the influence of White male data points on the joint distribution and expected moments. To address this issue, we created 1000 sub-sampled distributions while keeping the number of joint-states of predictive variables the same (i.e., the same number of Black women, White women, Black men, and White men). We then calculated the mixture distribution of all 1000 sub-sampled distributions and used that to perform the analysis. By resampling the data to enforce a maximum-entropy distribution, we are controlling for the problem of uneven sampling, such as oversampling done by the census-takers to ensure sufficient coverage.

While the resampling addresses some sampling biases (such as uneven numbers of one particular demographic group skewing estimates), there is also an issue of bias in the inference of the probability distributions and subsequent entropies themselves. Here we use the “naive” or maximum-likelihood estimator, where \( P(X = x) \) is estimated by the number of times \( x \) appears in the dataset, normalized by the total number of samples. The MLE estimator is known to have a persistent bias for small datasets (for review, see [20]), however given the very large number of samples in the data (on average \( 70718.9 \pm 8435.99 \) respondents in a given year), the effect of these biases is negligible. Furthermore, while bias corrections for entropy and mutual information are well-explored, in the context of the PID bias corrections are almost entirely unexplored, and so we leave this issue for future research.

By forcing a “flat” probability distribution on the space of joint identities, this can be understood as being a “causal” analysis in the sense of Pearl’s \( \text{do} \)-calculus model of
causality [51,52] and Woodward’s interventionist causal framework [53]. Specifically, we are leveraging the measure of “effective information” (first introduced by Tononi and Sporns [54]) which quantifies all of the causally-relevant relationships between two variables:

\[ EI(X \rightarrow Y) = I(X^{H_{\max}}; Y) \]  

(13)

where \( X^{H_{\max}} \) indicates that the probability distribution of the states of the predictor variable \( X \) are maximally entropic (in this case, uniform). The maximum entropy distribution is considered to be “causal” because it controls for any biases that might be introduced from the empirical distribution and preserves only the core “effective” relationships in the data. The consequence is that, by strictly forcing our distribution of identities to be uniform (i.e., \( P(\text{Race} = \text{Black} \land \text{Sex} = \text{Female}) = P(\text{Race} = \text{White} \land \text{Sex} = \text{Male}) \)), we are decomposing the effective information shared between identity and outcomes (as recorded in this data).

We compared three different sets of variables: \( I(\text{Race, Sex} ; \text{Income}) \), \( I(\text{Sex, Income} ; \text{Health}) \) and \( I(\text{Race, Sex, Income} ; \text{Health}) \), which represent a sampling of the interactions of three major identity groups (with income standing in as a proxy for class). We began by comparing the difference between the information in the “whole” (both identities considered together) to the sum of the “parts” (the information each individual identity added together), which serves as a useful heuristic for whether the relationship is dominated by intersectional synergy or not. Subsequently we did the full partial information decomposition to extract the relevant partial information-sharing modes. To compare between datasets, each value was normalized by the joint mutual information, giving the proportion of the total information (i.e., of all the information that race and sex provide about class, what proportion of that information is synergistic, redundant, unique, etc.).

Finally, we show that redundant and synergistic information make distinctly different predictions about the effects of identity, on outcomes, and discuss how these distinct modes of interaction can form a basis for modeling cross-identity solidarity and exclusive intersectional experiences, respectively.

2.6. Data & Code Availability

All data can be downloaded from the US Census Bureau website (https://www.census.gov/data/datasets/time-series/demo/cps/cps-asec.2020.html, accessed on 9 September 2021). All scripts necessary to recreate the analysis and figures are included as Supplementary Material. Data cleaning was done using the Pandas and Numpy packages and all information theoretic analysis was carried out in Python, using the Discrete Information Theory toolbox by James et al. [55]. Regression analysis was done using the Statsmodels package in Python [56].

3. Results

We began by examining the difference between the joint mutual information and the sum of the marginal entropies for each of our three relationships of interest across the whole decade (Figure 3). This is a practical heuristic sometimes called the “Whole Minus Sum” method [44,48] that quantifies the overall degree of redundancy or synergy-dominance in data. We found that, for all three relationships and across all ten years, the joint entropy was greater than the sum of the marginals, indicating that the relationship between various intersecting identities and target outcomes are consistently “greater than the sum of their parts,” or in the context of PID, “synergy-dominated”. The Whole Minus Sum heuristic provides a single, scalar measure that describes the interaction between multiple variables. In that respect, it is similar to standard linear regression approaches. If we compute a standard linear regression between race and sex on income, we see a single interaction effect (all coefficients significant at \( p < 10^{-10} \)).
Figure 3. The difference between the joint and the sum of the marginal mutual informations for three relationships. Across all ten years, for all three relationships assessed, the difference between the joint and the sum of the marginal mutual informations was consistently greater than 0, indicating that the joint state of all identity groups considered together disclosed more information about the outcomes (income, health status) than all the identities considered independently. The error-bars indicate a range of $\pm 1$ standard deviation.

While this tells us that there is some significant interaction between race and sex with respect to income, it tells us very little about the structure of that interaction. The coefficient is a single, scalar value and consequently cannot distinguish between the different kinds of interactions (redundant or synergistic). To determine the exact distributions of redundant, unique, and synergistic information, we then computed the full PIDs using the $I_{\text{min}}$ redundancy function (for visualization of the two triadic interactions, see Figure 4, for visualization of the results with $I_{\text{BROJA}}$ see Supplementary Material). We found that the relationship between race and sex on income had a high degree of both synergy and redundancy, with only limited unique information disclosed by only race or sex. On average, 51 $\pm$ 1.3% of the information disclosed by race and sex about income was synergistic in nature, while 42 $\pm$ 0.8% of the information was redundantly shared between the two predictor variables. The two unique components together disclosed only 7 $\pm$ 2% bit total. This decomposition into four different information-sharing modes shows that the general interaction revealed by the multiplicative interaction term in Table 1 is actually comprised of two different types of interaction, both present in roughly equal measure, while the particular implications of redundant and synergistic interactions in this context remains a question for further study, these results highlight how information decomposition-based approaches can reveal structure in data that linear regression is insensitive to.

We can also see that the particular structure of different interactions can vary widely depending on the particular variables being analyzed. In contrast to the interaction between race, sex, and income (where redundant and synergistic interactions dominated in roughly equal measure), the information that race and income provided about health status had a dramatically different structure: only 15.2 $\pm$ 2.2% of information were synergistic, and only 13.27 $\pm$ 1.8% was redundantly shared. Income uniquely disclosed 71.13 $\pm$ 3.88% of the information about health, with only 0.4 $\pm$ 0.82% being uniquely disclosed by race. Despite the clear differences in the breakdown of partial information, the joint mutual informations were very similar: $I(Race, Sex ; Income) = 0.057$ bit and $I(Race, Income ; Health) = 0.06$ bit—it is only when breaking down the information into its atomic components that we see meaningful differences. When considering the three-way effect of race, sex, and income on health, we found that 12.1 $\pm$ 1.9% of the total information was some form of redundant $(Red(Race, Sex, Income), Red(Race, Sex), Red(Race, Income), Red(Race, Sex))$, while another
15.48 ± 2.31% was some form of synergistic. Unique information account for 68.88 ± 3.76% of the total. Note that these do not sum up to 100% since there is a small amount of information distributed over more exotic and hard-to-interpret dependencies.

Figure 4. Partial information decomposition of $I(\text{Race, Sex}; \text{Income})$, $I(\text{Race, Income}; \text{Health Status})$ and $I(\text{Race, Sex, Income}; \text{Health Status})$. Top: The PID for race and sex on income. The information about income disclosed by race and sex is almost entirely either redundant or synergistic, race or sex individually disclose very little unique information. Middle: PID for race and income on health status. In contrast to the effects of race and sex, almost all information about health status is disclosed by income, although there is a non-trivial amount of both shared and synergistic information. This shows that interacting identities can have markedly different structures only revealed by information decomposition. Bottom: The PID for the relationship between race, sex, and income on health status. Due to the large number of distinct information-sharing modes, we aggregated all purely redundant terms, all purely unique terms, and all purely synergistic terms. The small difference between this total and the complete joint mutual information corresponds to exotic, higher-order interactions not reported here. We can see that, for all three relationships, the degree of informativeness remains remarkably constant over the decade, and that the overall degree of synergy, redundancy, and unique information is similar consistent.

Table 1. Income vs. Race and Sex. Least-squares regression results for the linear interaction between Race and Sex on Income.

| Variable         | $\beta$  | Std. Err. | T-Statistic | 95% C.I.     |
|------------------|---------|-----------|-------------|-------------|
| Intercept        | 1.6762 *| 0.007     | 238.521     | [1.662, 1.690]|
| Race             | -0.6375 *| 0.020    | -31.178     | [-0.678, -0.597]|
| Sex              | -0.6133 *| 0.010    | -62.519     | [-0.632, -0.594]|
| Race $\times$ Sex| 0.4405 *| 0.027     | 16.140      | [0.387, 0.494]|

| $R^2$             | 0.076   | No. of Observations | 59,858     |
|-------------------|---------|----------------------|-------------|
| F-Statistic       | 1641.0  | Prob > F             | 0.00        |
| AIC               | 1.832 x 10^5 | BIC                  | 1.832 x 10^5 |

* $p < 10^{-10}$.

When considering all of the years individually, it is clear that the informational relationships are stable over time (as evidenced by the comparatively low standard deviation terms). There does not appear to be a substantive increase, or decrease in how the various identities interact, suggesting that the intersectional relationships are consistent over
multiple years. When comparing the results of the $I_{\min}$ measure with the results of using the $I_{\text{BROJA}}$ measure, we found that the patterns were overwhelmingly similar: $I(\text{Race, Sex ; Income})$ was comprised largely of equal parts redundant and synergistic information, while $I(\text{Race, Income ; Health})$ decomposed into primarily information uniquely disclosed by income. Since $I_{\text{BROJA}}$ cannot accommodate more than two predictors, there was no way to calculate the joint-effects of race, sex, and income together on health. See Supplementary Material for visualization.

3.1. Untangling Redundant & Synergistic Information

The partial information decomposition framework describes two distinct ways that different identities can interact: they can both individually communicate the same information (redundancy) or they can jointly communicate information that is not not disclosed by any simpler combination of sources (synergy), which we identify with “intersectionality”. Following [31] and [57], we can relate the redundant and synergistic information to the logical disjunction (OR) and logical conjunction (AND), respectively. Consider the redundant information between two variables: it is information that is disclosed by either variable. An observer could choose one of the two variables at random, observe it and only it, and learn the same information as if they had chosen the other variable.

When considering the information redundantly present in both race and sex about income, we ask: how would we revise our estimate of a given person’s income if we knew that they were Black or a woman? Across all 1000 resampled distributions for the 2020 dataset, the overall expected income was $43,304.67, however if we restrict our analysis to only those individuals who are Black or a woman (excluding White men but including Black men, White women, and Black women), we find that the expected income drops to $39,006.65: this difference of $-4298.01 is the expected income penalty associated with having at least one of the identities in question, without knowing explicitly which one the person had. To consider the intersectional synergy, we can also ask what the expected income for someone who is both Black and a woman is: $35,908.84. This is a $7395.81 financial penalty specific to black women and another $3097.80 penalty that goes above and beyond being either Black or a woman. If we consider the Venn Diagram in Figure 2, the shared cost of being Black or a woman is represented by the innermost, hatched intersection, while the specific cost of being Black and a woman is represented by the outermost oval of synergy. This perspective is analogous to the definition of synergy given by Equation (9) and shows how the overall logic of the PID can be applied to other kinds of analysis relating “wholes” and “parts” but are not specifically operating on information directly. In the same way that synergistic information can be calculated as the difference between the whole and the union of the parts, the synergistic “cost” of misogyny can be calculated as the difference between the cost of both identities considered jointly and the simpler redundant cost shared redundantly.

Formally:

\[
\begin{align*}
\mathbb{E}[\text{Income}] &= 43,304.67, \\
\mathbb{E}[\text{Income}|\text{Black} \lor \text{Woman}] &= 39,006.65, \\
\mathbb{E}[\text{Income}|\text{Black} \land \text{Woman}] &= 35,908.84,
\end{align*}
\]

You can also do the same analysis considering privileged identities. Consider the logical disjunction of Whiteness and masculinity:

\[
\begin{align*}
\mathbb{E}[\text{Income}|\text{White} \lor \text{Man}] &= 45,775.02, \\
\mathbb{E}[\text{Income}|\text{White} \land \text{Man}] &= 56,212.66,
\end{align*}
\]

Being White or male comes with an additional benefit of $2,470.34, and the added benefit of being White and male above and beyond even that is $10,437.64, while being Black and woman comes with an intersectional “cost” of $\approx 3000$, being White and a
man comes with an intersectional “reward” of over $10,000. This shows that there are multiple ways that intersecting identities can interact, which we might call “redundant intersectionality” vs. “synergistic intersectionality”. The redundant intersectionality is the set of experiences or effects shared by both identities that is common to both of them. In contrast, the synergistic intersectionality is the extra effect specific to the identities in question that goes “above and beyond” the shared experiences of both independent identities. For visualization of this, see Figure 5. Considering specific realizations of all the variables is analogous to performing a local PID [57,58], which provides a finer-grained perspective than the the expected values, but with the same logical structure of redundant, unique, and synergistic dependencies. Given a suitable, local redundancy function (such as Finn and Lizier’s $I_Z$, Ince’s $I_{cs}$, or Makkeh et al.’s $I_{sx}$, the local PID could be done in terms of bit rather than USD.

![Expected Income Over Time](image1)

![Expected Income Over Time](image2)

**Figure 5.** The difference between the shared cost of marginalization, and the extra cost of synergistic intersectionality. Left: The expected income for the whole population over time, the expected income for that subset of the population that is Black or a woman, and, and the expect income for Black women. We can see that there is an extra cost incurred by being Black and a woman that is “above and beyond” the cost incurred by being Black or a woman. Right: the same plot, but considering Whiteness and masculinity. Once again, the joint relationship is strong different from the disjunction, only this time there is a synergistic benefit to being White and male, as opposed to a synergistic cost.

While the identity-specific decompositions described above can give us insights into how particular configurations of identities interact, each one only provides a small picture of the overall, intersectional dependencies. The local perspective does not tell us how race qua itself interacts with sex qua itself; it only gives us particular configurations of specific races and specific sexes. If we want to understand how race, as a concept that can be realized in many ways (White, Black, etc.), interacts with sex (which similarly can be realized in different ways), we need to abstract away from particular cases and consider the variables as objects in-and-of themselves. For example, the fact that there is a synergistic benefit accrued to White men does not automatically tell us anything about the interaction between race and sex more generally. The only way to get the full picture is via the partial information decomposition framework (which can then be localized if there is a particular interaction of interest that one wants to explore.)

**3.2. Robustness of the Pipeline**

We stress-tested our pipeline to assess how robust it might be to natural limitations in data collection: smaller sample sizes and noise in the data. By virtue of working with census data, we naturally have access to a larger-than-usual data set, although it was uncertain how well this kind of information-decomposition approach would work on
smaller data sets. To test this, we re-ran the whole inference for the $I(Race, Sex ; Income)$ analysis, using 20 different-sized subsets of the original data set each time. Subset sizes were arranged logarithmically, from 250 samples to the full-sized data set (59,858 samples). For each subset, we randomly sampled individual respondents, and replicated each subset 600 times, to create a distribution for each subsample size. We found that, for very small subset sizes, there was a significant 

over-estimation

of the difference between the joint mutual information and the sub of the marginals, although the estimates declined rapidly and converged towards the true value by $\approx 1000$ samples (see Figure 6 Top Left). When looking at the redundancy and synergy directly (normalized by the joint mutual information to give a proportion of the total information), we found that all subsample sizes found strong evidence of synergies between race and sex on income. The specific values changed, as did the relative ratio between them, but this is a promising result that, in general, the identification of intersectional synergies is possible even in a comparatively smaller dataset (see Figure 6 Bottom Left).

Figure 6. Robustness of the PID pipeline to noise and small sample sizes. The four plots show how the analytical pipeline described above performs when the initial dataset is shrunk, or has noise added to it. $I(Race, Sex ; Income)$ was used as an example data set, although comparable results can be seen for the other relationships. **Top Left** Assessing how the difference between the joint mutual information and the sum of the marginal mutual informations changes as the number of samples in the original data set is reduced. We can see that smaller samples tend to over-estimate the difference between whole - sum, although convergence happens reasonably quickly. **Top Right** Plotting the difference between the joint MI and the sum of the marginals as progressively more noise is added to the data. We can see that the result is somewhat sensitive to noise; with 10% noise in the data, the percentage change in whole-minus-sum information being $\approx -18.355\%$. **Bottom Left** The proportion of the total MI that is redundant or synergistic as the sample size increases, while the absolute values and relative ratios of the particular atoms change, the pipeline is always able to identify the existence of synergistic relationships in the data. **Bottom Right** The proportion of MI that is redundant or synergistic as the noise increases. Despite the change in whole-minus-sum values as noise increases, the relative ratios remain largely constant. These result show that the PID inference pipeline is reasonably robust both to decreased sample sizes and noise in the data.

To assess the effect of noise in the data, we took a similar approach. We re-ran our analysis, each time randomly permuting increasingly large subsets of the initial dataset.
The 20 subset sizes were arranged logarithmically from 0.01% to 10%, and each subset size was re-tested 600 times. We found that the whole-minus-sum heuristic showed a modest decrease as the randomization increased (see Figure 6 Top Right), however the percentage of the total mutual information that was redundant and synergistic remained extremely stable (see Figure 6 Bottom Right). These results collectively provide evidence that the PID analysis pipeline can be used for smaller, and noisier datasets while retaining sensitivity to the presence of higher-order statistical synergies.

3.3. Comparing PID and Linear Regression with Interaction Terms

As discussed in the Introduction, the most widely used methodology for assessing intersectionality in quantitative data is the use of multiplicative interaction terms in linear regression [19], however the linear interaction term fails to effectively disambiguate between different kinds of relationships between variables. To demonstrate this, we create two dummy datasets based on the empirical data (taken from the 2020 ASEC data), one of which is completely redundant, and another which is almost completely synergy-dominated, and show that the linear regressions (including interaction variables) are practically the same, despite the enforced differences in interaction structure. We argue that this indicates that regression-based approaches to intersectional data analysis have been missing important structures and relationships in the data that have clear implications for intersectional analysis. By constructing two datasets that are as distinct as possible in terms of underlying information structure, we aim to give the linear regression model the best possible chance of success: if the interaction term is, in some way, sensitive to redundant and synergistic interactions, then by constructing the largest possible difference, we are creating the strongest possible “signal” for the regression to pick up on.

We began with the relationship between Income and Health, known to be correlated in the real data. To create a triad dominated by redundancy, we created a dummy variable $D_{\text{red}}$ constructed in such a way that the information it contained about Health Status was completely redundant with Income:

$$D_{\text{Red}} = (1 + \text{Income}) \mod 4. \quad (14)$$

$D_{\text{red}}$ is functionally a copy of the Income dataset, with every value incremented by 1 (and wrapping around for values higher than the maximum income. Our joint mutual information decomposition then becomes:

$$I(\text{Income}, D_{\text{Red}}, \text{Health}). \quad (15)$$

When we do the full partial information decomposition using the Williams and Beer redundancy function $I_{\text{min}}$, we find that, as expected, all of the joint mutual information is redundant (the results are functionally the same when using $I_{\text{BROJA}}$, see Supplementary Material). This is an example of what we might might refer to as “source redundancy”: all of the redundant information is present in correlations between the source variables. Harder et al., proposed a distinction between “source redundancy” and “mechanistic redundancy”, the later of which is redundancy that is somehow intrinsic to the computation performed by the target and can be non-zero even in cases where both sources are independent. This distinction has been further explored by Goodwell and Kumar [36] and Ince et al. [59,60]. Untangling source and mechanistic redundancy remains an open problem in the field, and future work may show how it is reflected differently in different statistical analyses.

The second dummy variable was constructed to result in a synergy-dominated decomposition and is given by:

$$D_{\text{Syn}} = (\text{Income} + \text{Health}) \mod 2. \quad (16)$$

The operation is analogous to a generalized logical XOR operator (see [38] for reference). We decompose:
to produce a partial information decomposition that is 94.4% synergistic, with the remaining 5.6% of the information distributed over unique and redundant partial information terms (it is impossible to create a purely synergistic relationship involving Income and Health Status, since Income already discloses information about Health Status and they are therefore not independent), while these two distributions are unlikely to be representative of real-world datasets (we anticipate that almost all real-world data will combine elements of all partial information-sharing modes to some extent), by constructing simple test cases, we can get a better understanding of how the PID and the linear regression relate to each other. You can see the results of analyzing these two datasets into standard linear regressions with multiplicative interaction terms in Tables 2 and 3.

Table 2. Redundancy Dominated Synthetic Distribution. Least-squares regression results for the redundancy-dominated synthetic distribution. This distribution was generated using a circularly-shifted logical COPY operation.

| Variable   | β       | Std. Err. | T-Statistic | 95% C.I.       |
|------------|---------|-----------|-------------|----------------|
| Intercept  | 2.8980  | 0.025     | 116.486     | [2.849, 2.947] |
| Income     | −0.3359 | 0.009     | −39.427     | [−0.353, −0.319]|
| D_{Red}    | −0.3346 | 0.021     | −15.794     | [−0.376, −0.293]|
| Income × D_{Red} | 0.1336 | 0.010     | 13.467      | [0.114, 0.153]  |

R² 0.069  No. of Observations 59,858
F-Statistic 1485.0  Prob > F 0.00
AIC 1.682 × 10⁵  BIC 1.682 × 10⁵

* p < 10⁻¹⁰.

Table 3. Synergy-Dominated Synthetic Distribution. Least-squares regression results for the synergy-dominated synthetic distribution. This distribution was constructed using the generalized logical-XOR function.

| Variable   | β       | Std. Err. | T-Statistic | 95% C.I.       |
|------------|---------|-----------|-------------|----------------|
| Intercept  | 2.5841  | 0.009     | 291.454     | [2.567, 2.601] |
| Income     | −0.2934 | 0.005     | −60.137     | [−0.303, −0.284]|
| D_{Syn}    | −0.1645 | 0.012     | −13.554     | [−0.188, −0.141]|
| Income × D_{Syn} | 0.1473 | 0.007     | 21.213      | [0.134, 0.161]  |

R² 0.071  No. of Observations 59,858
F-Statistic 1520.0  Prob > F 0.00
AIC 1.681 × 10⁵  BIC 1.681 × 10⁵

* p < 10⁻¹⁰.

Looking at these regressions, it is clear that, despite the completely different information-structures, the linear regressions with multiplicative interaction terms finds very similar, statistically significant coefficients. Even a more involved analysis such as a Shapely decomposition would be unable to distinguish the “types” of interaction, instead treating it as a lump sum. These results can be visualized in Figure 7. This is a significant finding because it shows that linear regression with multiplicative interactions terms is incapable of distinguishing between “synergistic” intersectional relationships where the joint effects of identities are greater than the sum of their parts, and “redundant” interactions between identities that are not non-additive in the way originally discussed by Crenshaw and Bailey.
Figure 7. Linear regression with interaction terms fails to differentiate between synergy- and redundancy-dominated relationships Right: The partial information decomposition for two different dummy datasets. \(I(\text{Income}, D_{\text{red}}; \text{Health})\) was constructed to be completely redundant: Income and \(D_{\text{red}}\) disclose exactly the same information about Health. In contrast \(I(\text{Income}, D_{\text{syn}}; \text{Health})\) was constructed to be largely synergy dominated. Despite these clear differences in structure, the linear regressions with multiplicative interaction terms return essentially the same regression (see the surface plots on the right). This shows that linear regressions cannot uncover synergistic higher-order intersections of the sort that the intersectionality framework focuses on.

These are admittedly somewhat constructed dummy variables: no combinations of features in the census dataset displayed “pure” information structures, instead showing a mixture of unique, redundant, and synergistic dependencies that made comparisons of this type difficult. Instead, these results are strong evidence that linear regression-based statistical analyses cannot provide a full account of all the dependences in a set of variables. Given either of the coefficients presented in Table 2 or Table 3, it would be impossible to infer which one came from the redundancy-dominated or synergy-dominated configuration. Furthermore, even though both distributions were constructed de novo to have only multivariate dependences (redundancy or synergy) and no unique pairwise dependencies, both regressions returned significant single-variable regression coefficients. This suggests that, despite the lack of any unique information in the constructed distributions, the regression is assigning some dependency to the marginal variables. These results have clear implications for quantitative analysis of intersectionality in data. Prior studies that report significant multiplicative interaction coefficients as an indicator of intersectional relationships may, in fact, be conflating different kinds of interactions: redundant and synergistic dependencies may be indistinguishable to the regression. Furthermore, apparently “unique” effects (associated with the regression coefficients of the marginal variables), may not always correspond to the unique information disclosed by that variable in an information-theoretic sense.

4. Discussion

In this work, we have shown how the formal frameworks of information theory writ large, and partial information decomposition specifically, can be used to address the question of identifying intersectionality in quantitative data. Using a decade of data from the US Census Annual Social and Economic Supplement, we show that synergistic, “greater-than-the-sum-of-their-parts” interactions can be identified in the interactions between identities such as race, sex, and class on outcomes such as income and general health status and that the relative proportions of redundant, unique, and synergistic information remain stable across the decade. We furthermore show the most widely-used analyses, such as linear regression with interaction terms struggles to discriminate between synergistic
and redundant modes of information sharing, which has implications for how we might model the generative dynamics that create inequalities, as well as how we might effectively intervene to correct them.

This strongly suggests that salient social dynamics and relationships are being missed by many current approaches, and that introducing complementary frameworks alongside more standard analyses may be illuminating. This work fits into the broader project first outlined by Abbott in the seminal critique Transcending General Linear Reality [61]. Abbott argues that over-reliance on general linear models has influenced how researchers think about the world: that the map (consisting largely of linear relationships between interacting entities, assumptions about single effects being generated by single causes, etc.) is confused for the territory (the real world, which is highly non-linear and admits complex, potentially higher-order causal interactions). The results presented here show that the standard practices that make the assumptions Abbott describes may be missing potentially important relationships in empirical data: the existence of intersectional synergies are direct evidence of higher-order interactions between attributes that are both irreducible and, by virtue of the use of the effective information, suggestive of an effective relationship. While the evidence presented here indicates that linear interaction effects struggle to disambiguate redundant and synergistic higher-order interactions (as well as potentially presenting spurious unique information-sharing relationships), we should note that this is not the only way intersectionalities may be tested for in data using linear models. In addition to the above-mentioned Shapely decomposition, other analyses such as analysis of suppressor variables [62,63] may reveal higher-order dependencies. A rigorous, mathematical treatment linking PID and suppressor variables or Shapely decompositions may reveal previously un-recognized links between partial information decomposition and other methods.

Beyond technical and methodological advances, we feel that the conceptual distinction between “redundant” and “synergistic” interactions may be of theoretical interest. For example, some authors have suggested that intersectionality as a framework balkanizes individuals into small groups with limited shared solidarity. This concern is exemplified by by Naomi Zack, who writes:

“...as a theory of women’s identity, intersectionality is not inclusive insofar as members of specific intersections of race and class can create only their own feminisms” [64].

Zack goes on to argue:

“These ongoing segregations make it impossible for women to unite politically and they have not ended exclusion and discrimination among women, especially in the academy” [64].

Without weighing into political dimension of Zack’s argument, we claim that the partial information decomposition shows that “exclusive” intersections and “shared oppression” can co-exist in society. Continuing the example of misogynoiré as a relevant case, the partial information results suggests that a given individual who is both Black and a woman will simultaneously experience: synergistic costs associated with pure misogynoiré (corresponding to the synergistic part of the mutual information), “generic” anti-black racism and misogyny (corresponding to the unique part of the mutual information), as well as the “shared” costs experienced by all those who are either Black or a woman (including Black men and White women).

Different distributions of redundancy and synergy may reflect distinct generative dynamics and may have implications for our understanding of how inequalities emerge. For example, the synergy specific to Black women (“misogynoiré” [3]) represents what is generally theorized as “intersectionality” and reflects a particular configuration of social factors that single out Black women in particular (such as racist stereotypes, the particular intersection of medical racism and sexism, etc.). In contrast, the “shared cost” borne by everyone who is Black or a woman (including Black men and White women) likely reflects
a more general societal bias that the White man is the “default” American, and is not specific to any marginalized identity. Simply knowing that someone is Black or a woman (i.e., not White and male) is enough to prompt a downward revision of expected income. Untangling these different social dynamics may help elucidate the mechanisms by which different inequalities emerge, and inform how to most effectively rectify them.

Different information structures may also inform on the effectiveness of interventions: if the relationship between two identities and a target variable is unique, then only one variable likely needs to be intervened on, while if it is synergistic, then both predictor variables must be accounted for, as the effect is in the joint interaction that apparently drives the outcome. For example, if there is a cost to health uniquely disclosed by income (or financial class), then correcting that disparity should (ideally) require intervening just on financial status. In contrast, if the health disparity is synergistically predicted by both class and race (with little to no unique component), then intervening on individual financial situations may not work, as it is the joint-state of both that drives the disparity. If we consider, as another example, the result relating sex and race to income, the fact that there is a large synergistic effect, with little unique information disclosed by sex or race individually indicates that an attempt to improve income inequality will certainly have to wrangle with both identities simultaneously: a sex-only or race-only approach fails to account for the higher-order structures that generate the inequality.

Both of these claims are admittedly speculative at present, but stand as open questions of potential significance for future research on intersectionality.

The same analysis can be done with regards to the rewards of privilege. An interesting question for further research would be how privileged and marginalized identities interact: for conceptual simplicity we have focused on the two cases where identities “stack” in the same direction (Blackness and womanhood are both generally marginalized, while Whiteness and masculinity are generally privileged), although the synergies that may emerge when privileged and marginalized types of identities co-exist could yield interesting, novel insights. These complex interactions have been previously noted, and sometimes lead to apparently paradoxical or hard-to-understand patterns. For example, the Latino/Hispanic Health Paradox [65,66] describes the counter-intuitive finding that, despite facing an array of environmental, social, and economic challenges, Latino-Americans tend to have better health outcomes and live longer than the average American. Conversely, the finding that so-called “deaths of despair” are primarily (although not uniquely) associated with White men [67,68] suggests a complex interaction between ostensibly privileged identities and adverse outcomes such as suicide and drug addiction. By providing an analytical framework that allows for more fine-grained decomposition of interactions then is usually done, the PID framework may be able to help untangle some of these complexities.

We should stress that the framework presented here is not intended to either “prove”, or “disprove”, the existence of intersectionality as a theory. Instead, it is a statistical framework that can be used to identify irreducible intersectional synergies in large data sets. We hope that this may inform evidence-based discussion around social issues. The failure to find a particular synergistic relationship in a particular data set should not necessarily be grounds to claim that a particular intersection is “unreal” or “unimportant” given the restrictions inherent in working with limited data sets and the limitations of data collection. By the same token, the identification of an unexpected statistical synergy should be addressed critically and assessed in the context of previous scholarship.

Despite its utility, information theory comes with some particular limitations that must be addressed. The first is that the amount of data required for a reliable inference is much larger than what is required for a linear regression. This is particularly pressing for higher-order analyses that require reliable estimation of high-dimensional joint-probability spaces. For studies that rely on “big data” sets (e.g., Census data, data gathered from social media, etc.), this is unlikely to present a problem, although smaller-scale survey-based studies of local populations may not generate sufficient data, in which case linear methods may be preferable. Our stress-testing of our pipeline can allay those concerns somewhat,
although there are known lower-bounds on the ability to reliably infer probability distributions from finite datasets [20]. Another pressing issue is that the number of distinct information-sharing modes grows super-exponentially with the number of sources: for example, assessing the intersections of 5 distinct identities (e.g., race, sex/gender, class, sexuality, ability) on a single outcome (e.g., life-expectancy) would require computing 7,828,352 distinct values (many of which would be very hard to interpret higher-order terms such as the information about life expectancy disclosed by the joint state of race and sex/gender or class or the joint state of sexuality and ability). Given this difficulty, other frameworks for information decomposition may be worth exploring, such as the “synergy-first” proposal given by Quax et al. [69] or Rosas et al. [70].

Given these limitations, as well as the undeniable successes attributable to linear regression models, we do not suggest that regressions be tossed aside in favour of the PID exclusively, but rather that these different perspectives can be combined. Both frameworks have strengths and weaknesses, and we feel that there will be many contexts in which both together will provide insights that may not be accessible to either one alone. One possible way that the two methods could inform each-other is by using linear regression as a kind of filter. Given the tremendous number of factors that can be socially relevant to a single outcome, a complete PID analysis of every possible interaction would be unwieldy at best, and impossible at worst. Linear regression-based approaches, which are much more efficient to run, may be useful for identifying a restricted set of informative sources, which can then be decomposed using PID once an interesting interaction has been identified and more fine-grained analysis is required.

The relative simplicity of applying the partial information decomposition to the effective information opens the doors to a range of future research directions. For instance, it may be worthwhile to revisit previous empirical studies of intersectionality (e.g., those discussed in [19] to disambiguate which interactions are synergistic in nature, vs. which ones are redundant in nature. It also suggests that researchers doing a multiplicative-interaction based analysis of intersectional relationships may consider supplementing their analysis with a PID-based framework to explicitly untangle redundant and synergistic components. Alternately, being able to untangle what features are redundantly predictive of an outcome, vs. which are synergistic can provide deeper insights into the generative dynamics of social relationships. Finally, recent work on generalizing the PID framework to accommodate multiple targets [71] can also expand the space of accessible interactions and further untangle complex dependencies between identities and outcomes (for example, how do race and sex joint inform on income and health together). We anticipate that, as work on PID is further developed, a large space of hitherto-unrecognized, higher-order interactions (both redundant and synergistic) will come into view and provide a rich ground on which to build a deeper understanding of complex interactions in society and health.

5. Conclusions

Despite the extensive and valuable work that has been done integrating intersectional frameworks in data analysis and social science research, the most commonly-used methods (e.g., linear regression with multiplicative interaction terms) fail to capture the range of ways that different identities can interact and predict outcomes. Information theory provides an appealing alternative framework that allows users to identify different modes of interactions of intersecting identities, while simultaneously shedding the requirements imposed by assumptions of linearity. These modes, which we called “redundant intersections” and “synergistic intersections” correspond to shared costs/benefits of multiple interacting identities and the exclusive costs/benefits, respectively. Examples can be seen in the redundant and synergistic interactions between race and sex considered jointly, allowing the identification of a strong synergistic cost of misogynoir, as well as a shared common cost between Blackness and womanhood. We hope that the framework detailed here will enable novel and insightful work on intersectionality in empirical data.
Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/e24101387/s1, Figure S1: The PIDs for $I(\text{Race, Sex ; Income})$ and $I(\text{Race, Income ; Health})$ using the BROJA function. [30,47,48]. Figure S2: The PIDs for the synergy and redundancy dominated synthetic datasets using the BROJA function. Scripts.zip: the Python scripts required to replicate the analysis.

Author Contributions: T.F.V. and P.K. conceptualized the project. T.F.V. performed data analysis, drafted, and edited the manuscript. P.K. assisted with interpretation and editing the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: T.F.V. and P.K. are supported by the NSF-NRT grant 1735095, Interdisciplinary Training in Complex Networks and Systems at Indiana University Bloomington.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: All of the data are available on the US Census website: https://www.census.gov/data/datasets/time-series/demo/cps/cps-asec.html (accessed on 9 September 2021).

Acknowledgments: Brea Perry has given us crucial input which made the paper significantly better. We would like to thank Byungkyu Lee and Michael Schultz for their detailed feedback and discussion of the early versions of this manuscript. We would also like to thank Zackary Dunivin, and Bradi Heaberlin for thoughtful discussions about intersectionality, statistics, and the concept of this project. Finally, we would like to thank Olaf Sporns.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Crenshaw, K. Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics. *Univ. Chic. Leg. Forum* 1989, 1989, 8.
2. Crenshaw, K. Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color. *Stanf. Law Rev.* 1991, 43, 1241–1299. [CrossRef]
3. Bailey, M. *Misogynoir Transformed: Black Women’s Digital Resistance*; NYU Press: New York, NY, USA, 2021;
4. Cho, S.; Crenshaw, K.W.; McCall, L. Toward a Field of Intersectionality Studies: Theory, Applications, and Praxis. *Signs J. Women Cult. Soc.* 2013, 38, 785–810. [CrossRef]
5. Taylor, Y.; Hines, S.; Casey, M. *Theorizing Intersectionality and Sexuality*; Springer: Berlin, Germany, 2010.
6. Ferlatte, O.; Salway, T.; Hankivsky, O.; Trussler, T.; Oliffe, J.L.; Marchand, R. Recent Suicide Attempts Across Multiple Social Identities Among Gay and Bisexual Men: An Intersectionality Analysis. *J. Homosex.* 2018, 65, 1507–1526. [CrossRef] [PubMed]
7. Block, D.; Corona, V. Exploring class-based intersectionality. *Lang. Cult. Curric.* 2014, 27, 27–42. [CrossRef]
8. Nirmala, E.; Minear, A. Unspeakingly Offenses: Untangling Race and Disability in Discourses of Intersectionality. *J. Lit. Cult. Disabil. Stud.* 2010, 4, 127–146.
9. Ben-Moshe, L.; Magaña, S. An Introduction to Race, Gender, and Disability: Intersectionality, Disability Studies, and Families of Color. *Women Gand. Fam. Color* 2014, 2, 105–114. [CrossRef]
10. Erez, E.; Harper, S. Intersectionality, Immigration, and Domestic Violence. In *The Handbook of Race, Ethnicity, Crime, and Justice*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2018; pp. 457–474. [CrossRef]
11. Romero, M. The inclusion of citizen ship intersectionality: What immigration raids tell us about mixed-status families, the state, and assimilation. *Int. J. Social. Fam.* 2008, 34, 131–152.
12. Viruell-Fuentes, E.A.; Miranda, P.Y.; Abdulrahim, S. More than culture: Structural racism, intersectionality theory, and immigrant health. *Soc. Sci. Med.* 2012, 75, 2099–2106. [CrossRef]
13. Yuval-Davis, N. Intersectionality, Citizenship and Contemporary Politics of Belonging. *Crit. Rev. Int. Soc. Political Philos.* 2007, 10, 561–574. [CrossRef]
14. Hankivsky, O.; Cormier, R. Intersectionality and Public Policy: Some Lessons from Existing Models. *Political Res. Q.* 2011, 64, 217–229. [CrossRef]
15. Hankivsky, O.; Jordan-Zachery, J.S. Introduction: Bringing Intersectionality to Public Policy. In *The Palgrave Handbook of Intersectionality in Public Policy*; Hankivsky, O., Jordan-Zachery, J.S., Eds.; The Politics of Intersectionality; Springer International Publishing: Cham, Switzerland, 2019; pp. 1–28. [CrossRef]
16. Garcia, T.C.; Zajicek, A. Incorporating Intersectionality in Public Policy: A Systematic Literature Review. *Humanit. Soc.* 2021, 46. [CrossRef]
17. Bowleg, L. When Black + Lesbian + Woman ≠ Black Lesbian Woman: The Methodological Challenges of Qualitative and Quantitative Intersectionality Research. *Sex Roles* 2008, 59, 312–325. [CrossRef]
18. Scott, N.A.; Siltanen, J. Intersectionality and quantitative methods: Assessing regression from a feminist perspective. *Int. J. Soc. Res. Methodol.* 2017, 20, 373–385. [CrossRef]
19. Rouhani, S. Intersectionality-informed Quantitative Research: A Primer. *Am. J. Public Health* 2014, 103, 1082.
20. Bosomtwe, T.; Barnett, L.; Harré, M.; Lizier, J.T. *An Introduction to Transfer Entropy: Information Flow in Complex Systems*; Springer: Berlin, Germany, 2016.
21. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; John Wiley & Sons: Hoboken, NJ, USA, 2012.
22. Drăgulescu, A.; Yakovenko, V.M. Exponential and power-law probability distributions of wealth and income in the United Kingdom and the United States. *Phys. A Stat. Mech. Its Appl.* 2001, 299, 213–221. [CrossRef]
23. Nair, J.; Wiernik, A.; Zwart, B. *The Fundamentals of Heavy Tails: Properties, Estimation, and Estimation*; Cambridge University Press: Cambridge, UK, 2022.
24. Barabási, A.L. The origin of bursts and heavy tails in human dynamics. *Nature* 2005, 435, 207–211. [CrossRef]
25. Mize, T.D. Best Practices for Estimating, Interpreting, and Presenting Nonlinear Interaction Effects. *Sociol. Sci.* 2019, 6, 81–117. [CrossRef]
26. Cliff, O.M.; Novelli, L.; Fulcher, B.D.; Shine, J.M.; Lizier, J.T. Assessing the significance of directed and multivariate measures of linear dependence between time series. *Phys. Rev. Res.* 2021, 3, 013145. [CrossRef]
27. Barnett, L.; Barnett, A.B.; Seth, A.K. Granger Causality and Transfer Entropy Are Equivalent for Gaussian Variables. *Phys. Rev. Lett.* 2009, 103, 238701. [CrossRef]
28. Kraskov, A.; Stögbauer, H.; Grassberger, P. Estimating Mutual Information. *Phys. Rev. E* 2004, 69, 066138. [CrossRef] [PubMed]
29. Lizier, J.T. JIDT: An Information-Theoretic Toolkit for Studying the Dynamics of Complex Systems. *Front. Robot. AI* 2014, 1, 11. [CrossRef]
30. Williams, P.L.; Beer, R.D. Nonnegative Decomposition of Multivariate Information. *arXiv* 2010, arXiv:1004.2515.
31. Gutknecht, A.J.; Vibral, M.; Makkeh, A. Bits and Pieces: Understanding Information Decomposition from Part-whole Relationships and Formal Logic. *arXiv* 2020, arXiv:2008.09535.
32. Tax, T.M.S.; Mediano, P.A.M.; Shanahan, M. The Partial Information Decomposition of Generative Neural Network Models. *Entropy* 2017, 19, 474. [CrossRef]
33. Colonier, N.; Van de Steen, F.; Uddin, L.Q.; Pollock, R.A.; Calhoun, V.D.; Marinazzo, D. Disambiguating the role of blood flow and global signal with partial information decomposition. *Neuroimage* 2020, 213, 116699. [CrossRef]
34. Timme, N.M.; Ito, S.; Myoshiyunchenoko, M.; Nigam, S.; Shimono, M.; Yeh, F.C.; Hotowy, P.; Litke, A.M.; Beggs, J.M. High-Degree Neurons Feed Cortical Computations. *PLoS Comput. Biol.* 2016, 12, e1004858. [CrossRef]
35. Newman, E.L.; Varley, T.F.; Parakkattu, V.K.; Sherrill, S.P.; Beggs, J.M. Revealing the Dynamics of Neural Information Processing with Multivariate Information Decomposition. *Entropy* 2022, 24, 930. [CrossRef]
36. Goodwell, A.E.; Kumar, P. Temporal information partitioning: Characterizing synergy, uniqueness, and redundancy in interacting environmental variables. *Water Resour. Res.* 2017, 53, 5920–5942. [CrossRef]
37. Luppi, A.I.; Mediano, P.A.M.; Rosas, F.E.; Allanson, J.; Pickard, J.D.; Carhart-Harris, R.L.; Williams, G.B.; Craig, M.M.; Finoia, P.; Owen, A.M.; et al. A Synergistic Workspace for Human Consciousness Revealed by Integrated Information Decomposition. *bioRxiv* 2020. [CrossRef]
38. Rosas, F.E.; Mediano, P.A.M.; Jensen, H.J.; Seth, A.K.; Barrett, A.B.; Carhart-Harris, R.L.; Bor, D. Reconciling emergences: An information-theoretic approach to identify causal emergence in multivariate data. *PLoS Comput. Biol.* 2020, 16, e1008289. [CrossRef] [PubMed]
39. Varley, T.F.; Hoel, E. Emergence as the conversion of information: A unifying theory. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* 2022, 380, 20210150. [CrossRef] [PubMed]
40. Varley, T.F. Flickering emergences: The question of locality in information-theoretic approaches to emergence. *arXiv* 2022, arXiv:2208.14502.
41. Luppi, A.I.; Mediano, P.A.M.; Rosas, F.E.; Holland, N.; Fryer, T.D.; O’Brien, J.T.; Rowe, J.B.; Menon, D.K.; Bor, D.; Stamatakis, E.A. A synergistic core for human brain evolution and cognition. *bioRxiv* 2020. [CrossRef] [PubMed]
42. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* 1948, 27, 379–423. [CrossRef]
43. Lizier, J.T. *The Local Information Dynamics of Distributed Computation in Complex Systems*; Springer: Berlin/Heidelberg, Germany, 2013. [CrossRef]
44. Baldazzi, D.; Tononi, G. Integrated information in discrete dynamical systems: Motivation and theoretical framework. *PLoS Comput. Biol.* 2008, 4, e1000091. [CrossRef]
45. Rosas, F.; Mediano, P.A.M.; Gastpar, M.; Jensen, H.J. Quantifying High-order Interdependencies via Multivariate Extensions of the Mutual Information. *Phys. Rev. E* 2019, 100, 032305. [CrossRef]
46. Chechik, G.; Globerson, A.; Anderson, M.; Young, E.; Nelken, I.; Tishby, N. Group Redundancy Measures Reveal Redundancy Reduction in the Auditory Pathway. In *Proceedings of the Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2001; Volume 14.
47. Bertschinger, N.; Rauh, J.; Olbrich, E.; Jost, J.; Ay, N. Quantifying Unique Information. *Entropy* 2014, 16, 2161–2183. [CrossRef]
48. Griffith, V.; Koch, C. Quantifying synergistic mutual information. *arXiv* 2014. arXiv:1205.4265.
49. Bertschinger, N.; Rauh, J.; Olbrich, E.; Jost, J. Shared Information—New Insights and Problems in Decomposing Information in Complex Systems. *arXiv* 2013, arXiv:1210.5902.
50. Faber, S.P.; Timme, N.M.; Beggs, J.M.; Newman, E.L. Computation is concentrated in rich clubs of local cortical networks. *Netw. Neurosci.* 2018, 3, 384–404. [CrossRef]
51. Pearl, J. Causal Inference. In Proceedings of the Workshop on Causality: Objectives and Assessment at NIPS 2008, Whistler, Canada, 12 December 2008.

52. Pearl, J.; Glymour, M.; Jewell, N.P. Causal Inference in Statistics: A Primer; John Wiley & Sons: Hoboken, NJ, USA, 2016.

53. Woodward, J. Making Things Happen: A Theory of Causal Explanation; Oxford Studies in Philosophy of Science; Oxford University Press: Oxford, UK; New York, NY, USA, 2005.

54. Tononi, G.; Sporns, O. Measuring information integration. BMC Neurosci. 2003, 4, 31. [CrossRef]

55. James, R.; Ellison, C.; Crutchfield, J. dit: A Python package for discrete information theory. J. Open Source Softw. 2018, 3, 738. [CrossRef]

56. Seabold, S.; Perktold, J. Statsmodels: Econometric and Statistical Modeling with Python. In Proceedings of the 9th Python in Science Conference (SciPy 2010), Austin, TX, USA, 3–28 July 2010; pp. 92–96. [CrossRef]

57. Makkeh, A.; Gutknecht, A.J.; Wibral, M. Introducing a differentiable measure of pointwise shared information. Phys. Rev. E 2021, 103, 032149. [CrossRef]

58. Finn, C.; Lizier, J.T. Pointwise Partial Information Decomposition Using the Specificity and Ambiguity Lattices. Entropy 2018, 20, 297. [CrossRef] [PubMed]

59. Ince, R.A.A. The Partial Entropy Decomposition: Decomposing multivariate entropy and mutual information via pointwise common surprisal. arXiv 2017, arXiv: 1702.01591.

60. Kay, J.W.; Ince, R.A.A.; Dering, B.; Phillips, W.A. Partial and Entropic Information Decompositions of a Neuronal Modulatory Interaction. Entropy 2017, 19, 560. [CrossRef]

61. Abbott, A. Transcending General Linear Reality. Sociol. Theory 1988, 6, 169. [CrossRef]

62. Horst, P. The Prediction of Personal Adjustment: A Survey of Logical Problems and Research Techniques, with Illustrative Application to Problems of Vocational Selection, School Success, Marriage, and Crime; Social Science Research Council: New York, NY, USA, 1941. [CrossRef]

63. Friedman, L.; Wall, M. Graphical Views of Suppression and Multicollinearity in Multiple Linear Regression. Am. Stat. 2005, 59, 127–136. [CrossRef]

64. Zack, N. Inclusive Feminism: A Third Wave Theory of Women’s Commonality; Rowman & Littlefield Publishers: Lanham, MD, USA, 2005.

65. Acevedo-Garcia, D.; Bates, L.M. Latino Health Paradoxes: Empirical Evidence, Explanations, Future Research, and Implications. In Latinas/os in the United States: Changing the Face of América; Rodriguez, H., Sáenz, R., Menjívar, C., Eds.; Springer: Boston, MA, USA, 2008; pp. 101–113. [CrossRef]

66. Ruiz, J.M.; Hamann, H.A.; Mehl, M.R.; O’Connor, M.F. The Hispanic health paradox: From epidemiological phenomenon to contribution opportunities for psychological science. Group Process. Intergroup Relat. 2016, 19, 462–476. [CrossRef]

67. Scutchfield, F.D.; Keck, C.W. Deaths of Despair: Why? What to Do? Am. J. Public Health 2017, 107, 1564–1565. [CrossRef] [PubMed]

68. Tilstra, A.M.; Simon, D.H.; Masters, R.K. Trends in “Deaths of Despair” Among Working-Aged White and Black Americans, 1990–2017. Am. J. Epidemiol. 2021, 190, 1751–1759. [CrossRef] [PubMed]

69. Quax, R.; Har-Shemesh, O.; Sloot, P.M.A. Quantifying Synergistic Information Using Intermediate Stochastic Variables. Entropy 2017, 19, 85. [CrossRef]

70. Rosas, F.E.; Mediano, P.A.M.; Rassouli, B.; Barrett, A.B. An operational information decomposition via synergistic disclosure. J. Phys. A Math. Theor. 2020, 53, 485001. [CrossRef]

71. Varley, T.F. Decomposing past and future: Integrated information decomposition based on shared probability mass exclusions. arXiv 2022, arXiv:2202.12992.