PROCEEDINGS A

royalsocietypublishing.org/journal/rspa

Research

Cite this article: Gutknecht AJ, Wibral M, Makkeh A. 2021 Bits and pieces: understanding information decomposition from part-whole relationships and formal logic. *Proc. R. Soc. A* **477**: 20210110.
https://doi.org/10.1098/rspa.2021.0110

Received: 4 March 2021
Accepted: 10 June 2021

Subject Areas:
statistics, mathematical logic, computational biology

Keywords:
information theory, partial information decomposition, part-whole relations, logic, multivariate statistical dependency, neural networks

Author for correspondence:
A. J. Gutknecht
e-mail: agutkne@uni-goettingen.de

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.5490672.

THE ROYAL SOCIETY PUBLISHING

Bits and pieces: understanding information decomposition from part-whole relationships and formal logic

A. J. Gutknecht[1,2], M. Wibral[1] and A. Makkeh[1]

[1]Campus Institute for Dynamics of Biological Networks, Georg-August University, Goettingen, Germany
[2]MEG Unit, Brain Imaging Center, Goethe University, Frankfurt, Germany

AJG, 0000-0002-2704-6944; AM, 0000-0002-3581-8262

Partial information decomposition (PID) seeks to decompose the multivariate mutual information that a set of source variables contains about a target variable into basic pieces, the so-called 'atoms of information'. Each atom describes a distinct way in which the sources may contain information about the target. For instance, some information may be contained uniquely in a particular source, some information may be shared by multiple sources and some information may only become accessible synergistically if multiple sources are combined. In this paper, we show that the entire theory of PID can be derived, firstly, from considerations of part-whole relationships between information atoms and mutual information terms, and secondly, based on a hierarchy of logical constraints describing how a given information atom can be accessed. In this way, the idea of a PID is developed on the basis of two of the most elementary relationships in nature: the part-whole relationship and the relation of logical implication. This unifying perspective provides insights into pressing questions in the field such as the possibility of constructing a PID based on concepts other than redundant information in the general n-sources case. Additionally, it admits of a particularly accessible exposition of PID theory.

1. Introduction

Partial information decomposition (PID) is an example of a rare class of problems where a deceptively simple question has perplexed researchers for many years, leading to heated disputes over possible solutions [1], simple but incomplete answers [2], and even to statements that the question should not be asked [3].

The core question of PID is how the information carried by multiple source variables about a target variable is distributed over the source variables. In other words, it is the information theoretic question of 'who knows what about the target variable'. Intuitively, answering this question involves finding out which information we could get from multiple variables alike (called redundant or shared information), which information we could get only from specific variables, but not the others (called unique information), and which information we can only obtain when looking at some variables together (called synergistic information).

Examples of questions involving PID are found in almost all fields of quantitative research. In neuroscience, for instance, we are interested in how the activity of multiple neurons, that were recorded in response to a stimulus, can provide information about (i.e. encode) the stimulus. Specifically, we are interested in whether the information provided by those neurons about the stimulus is provided redundantly, such that we can obtain it from many (or any) of the recorded neural responses, or whether certain aspects are only present uniquely in individual neurons but not others; finally, we may find that we need to analyse all neural responses together to decode the stimulus—a case of synergy. All three ways of providing information about the stimulus may coexist and the aim of PID analysis is to determine to what degree each of them is present [4].

In this way, PID can be used as a framework for systematically testing and comparing theories of neural processing (such as predictive coding [5] or coherent infomax [6]) in terms of their information theoretic 'footprint', i.e. in terms of the amounts of unique, redundant or synergistic information processing predicted by the theory. The key idea is to identify such theories with a specific information theoretic goal function (e.g. maximize redundancy while at the same time allowing for a certain degree of unique information). One may then investigate empirically whether a given neural circuit in fact maximizes the goal function in question or one may use the PID framework to come up with entirely new goal functions [7].

The PID problem also arises in cryptography in the context of so-called 'secret sharing' [8]. The idea is that a multiple participants (the sources) each hold some partial information about a particular piece of information called the secret (the target). However, the secret can only be accessed if certain participants combine their information. In this context, PID describes how access to the secret is distributed over the participants.

The PID framework has furthermore been used to operationalize several core concepts in the study of complex and computational systems. These concepts include for instance the notion of information modification [9,10] which has been suggested along with information storage and transfer as one of three fundamental component processes of distributed computation. It has also been proposed that the concepts of emergence and self-organization can be made quantifiable within the PID framework [11,12].

Despite the universality of the PID problem, solutions have only arisen very recently, and the work on consolidating and on distilling them into a coherent structure is still in progress. In this paper, we aim to do so by rederiving the theory of PID from the perspective of mereology (the study of parthood relations) and formal logic. The general structure of PID arrived at in this way is equivalent to the one originally described by Williams & Beer [13]. However, our derivation has the advantage of tackling the problem directly from the perspective of the *parts* into which the information carried by the sources about the target is decomposed, the so-called 'atoms of information'. By contrast, the formulation used until now takes an indirect approach via the concept of redundant information. Furthermore, the approach described here is based on particularly elementary concepts: parthood between information contributions and logical implication between statements about source realizations.

The remainder of this paper is structured as follows. First, in §2, we derive the general structure underlying PID from considerations of elementary parthood relationships between information contributions. This structure is general in the sense that it still leaves open the possibility for multiple alternative measures of information decomposition. We show that the axioms underlying the formulation by Williams & Beer [13,14] can be proven within the framework described here. In §3, we use formal logic to derive a specific PID measure and in this way provide a complete solution to the information decomposition problem. Section 4 shows that there is an intriguing connection between formal logic and PID in that the mathematical lattice structure underlying information decomposition is isomorphic to a lattice of logical statements ordered by logical implication. This gives rise to a completely independent exposition of PID theory in terms of a hierarchy of logical constraints on how information about the target can be accessed. In §5, we show that the ideas presented here can be used to systematically answer the question of whether a (full n-sources) PID can be induced by measures other than redundant information such as synergy or unique information. Before concluding in §7, we briefly address the important distinction between parthood relations and quantitative relations in §6.

2. The parthood perspective

Suppose there are n source variables $S_1, \ldots, S_n$ carrying some joint mutual information $I(T : S_1, \ldots, S_n)$ [15,16] about some target variable T (see figure 1*a*). The goal of PID is to decompose this joint mutual information into its component parts, the so-called *atoms* of information. As explained in the §1, these parts are supposed to represent unique, redundant, and synergistic information contributions. Now, what distinguishes these contributions are their defining part-whole relationships to the information provided by the different source variables: the information uniquely associated with one of the sources in only part of the information provided by *that* source and not part of the information provided by any other source. The information provided redundantly by multiple sources is part of the information carried by *each* of these sources. And the information provided synergistically by multiple source is only part of the information carried by them jointly but not part of the information carried by any of them individually. For this reason, it seems natural to make the part-whole relationship between pieces of information the basic concept of PID. The goal of this section is to make this idea precise, and in this way, to open up a new perspective for thinking about PID.

The underlying idea is that any theory should be put on the foundation of as simple and elementary concepts as possible. The part-whole relation is one of the most basic relationships in nature. It appears on all spatial and temporal scales: atoms are parts of molecules, planets are parts of solar systems, the phase of hyperpolarization is part of an action potential, infancy is part of a human beings life. Moreover, it is not a purely scientific concept but is also ubiquitous in ordinary life: we say, for instance, that a prime minister is part of the government or that a slice of pizza is part of the whole pizza. This ubiquity makes it particularly easy to think in terms of part-whole relationships. We hope, therefore, that starting from this vantage point will provide a particularly accessible and intuitive exposition of PID. This factor is of particular importance when it comes to the practical application of PID to specific scientific questions and the interpretation of the results of a PID analysis.

Developing the theory of PID means that we have to answer three questions:

(i) What are atoms of the decomposition supposed to mean, i.e. what *type* of information should they represent?
(ii) How many atoms are there for a given number of information sources?
(iii) How large are the different atoms of information given a specific joint probability distribution of sources and target? How many *bits* of information does each atom provide?

In the following sections, we will tackle each of these questions in turn.

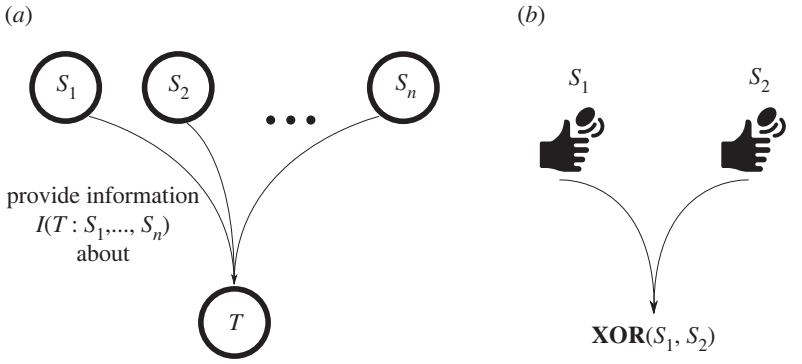

Figure 1. (*a*) The general partial information decomposition problem is to decompose the joint mutual information provided by n source variables $S_1, \ldots, S_n$ about a target variables T into its component parts. (*b*) Illustration of the exclusive-or example. The sources are two independent coin flips. The target is 0 just in case both coins come up heads or both come up tails. It is 1 if one of the coins is heads while the other is tails. Coin tossing icons made by Freepik, www.flaticon.com.

(a) What do the atoms of information mean?

Asking how to decompose the joint mutual information into its component parts is a bit like asking 'How to slice a cake?'. Of course, there are many possible ways to do so, and, hence, there is no unique answer to the question. In order to make the question more precise, we first have to provide a criterion according to which we would like to decompose the joint mutual information. This is what this section is about. What are the atoms of information supposed to mean in the end, i.e. what *type* of information do they represent?

To a first approximation, the core idea underlying the parthood approach to PID is to decompose the joint mutual information $I(T : S_1, \ldots, S_n)$ into information atoms, such that each atom is characterized by its parthood relations to the mutual information provided by the different sources. For instance, one atom of information will describe that part of the joint mutual information which is part of the information provided by *each* source, i.e. the information that is redundant to all sources. Another atom will describe the part of the joint mutual information that is only part of the information provided by the first source, i.e. it is unique to the first source. And so on.

Now, we have to refine this idea a bit: it is important to realize that it would not be enough to consider parthood relations to information provided by *individual* sources. The reason is that a *collection* of sources may provide some information that is not contained in any individual source but which only arises by *combining* the information from multiple sources in that collection. The classical example for this phenomenon is the logical exclusive-or shown in figure 1, right. In this example, the sources are two independent coin flips. The target is the exclusive-or of the sources, i.e. the target is 0 just in case both coins come up heads or both come up tails, and it is 1 otherwise. Initially, the odds for the target being zero or one respectively are 1:1 because there are four equally likely outcomes in two of which the target is 1 while it is 0 in the other two. Now, if we are told the value of one of the coins, these odds are not affected, and accordingly, we do not obtain any information about the target. For instance, if we are told that the first coin came up heads there are two equally likely outcomes left: Heads-Heads and Heads-Tails. In the first case, the target is zero and in the second case it is one. Hence, the odds are still 1:1. On the other hand, if we are told the value of *both* coins, then we *know* what the value of the target is. In other words, we obtain complete information about the target.

There are two conclusions to be drawn from examples like this:

(i) There are cases in which multiple information sources combined provide some information that is not contained in any individual source. This type of information is generally called *synergistic information*.

Table 1. Parthood table for the case of two information sources. Each row characterizes a particular atom of information in terms of its parthood relationships with the mutual information provided by the different collections of sources. The bold entries are enforced by the constraints that there is no information in the empty collection of sources and any piece of information is part of the information carried by the full set of sources about the target.

part of	{}	{1}	{2}	{1, 2}
Π_1 (synergy)	**0**	0	0	**1**
Π_2 (unique)	**0**	1	0	**1**
Π_3 (unique)	**0**	0	1	**1**
Π_4 (shared)	**0**	1	1	**1**

(ii) Any reasonable theory of information should be compatible with the existence of synergistic information. In particular, it should allow that, in some cases, the information provided jointly by multiple sources is larger than the sum of the individual information contributions provided by the sources.

Regarding the second point we may note that classical information theory satisfies this constraint because in some cases

$$I(T : S_1, S_2) > I(T : S_1) + I(T : S_2). \tag{2.1}$$

In fact, in the exclusive-or example, each individual source provides zero bits of information while the sources combined provide one bit of information.

Based on these consideration we may rephrase the basic idea of the parthood approach as: we are looking for a decomposition of the joint mutual information into atoms such that each atom is characterized by its parthood relations to the information carried by the different possible *collections* of sources about the target. Of course, we allow collections containing only a single source, such as {1}, as a special case. Note that we will generally refer to source variables and collections thereof *by their indices*. So instead of writing $\{S_1\}$ and $\{S_1, S_2\}$ to refer to the first source and the collection containing the first and second source, we write {1} and {1, 2}, respectively. There are several important technical reasons for this that will become apparent in the following sections. For now, it is sufficient to just think of it as a shorthand notation.

Let us now investigate how the idea of characterizing the information atoms by parthood relations plays out in the simple case of two sources S_1 and S_2. In this case, there are four collections:

(i) The empty collection of sources {}
(ii) The collection containing only the first source {1}
(iii) The collection containing only the second source {2}
(iv) The collection containing both sources {1, 2}

Now, in order to characterize an information atom Π we have to ask for each collection **a**: Is Π part of the information provided by **a**? For two of the collections we can answer this question immediately for all Π: First, no atom of information should be contained in the information provided by the empty collection of sources because there is no information in the empty set. If we do not know any source, then we cannot obtain any information from the sources. Second, any atom of information should be contained in the mutual information provided by the full set of sources since this is precisely what we want to decompose into its component parts. Regarding collections {1} and {2} we are free to answer yes or no leaving four possibilities as shown in table 1.

The first possibility (first row of table 1) is an atom of information that is only part of the information provided by the sources jointly but not part of the information in either of the individual sources. This is the *synergistic information*. The second possibility (second row) is an atom that is part of the information provided by the first source but which is not part of the

Table 2. Example of Boolean function that is not a parthood distribution. Bold entries violate the monotonicity constraint.

part of	{}	{1}	{2}	{3}	{1, 2}	{1, 3}	{2, 3}	{1, 2, 3}
	0	1	0	0	**0**	**0**	0	1

information in the second source. This atom of information describes the *unique information* of the first source. Similarly, the third possibility (third row) is an atom describing information uniquely contained in the second source. The fourth and last possibility (fourth row) is an atom that is part of the information provided by *each* source. This is the information *redundantly provided* or *shared by* the two sources.

So based on considerations of parthood we arrived at the conclusion that there should be exactly four atoms of information in the case of two source variables. Each atom is characterized by its parthood relations to the mutual information provided by the different collections of sources. These relationships are described by the rows of table 1 which we will call *parthood distributions*. Each atom Π is formally represented by its parthood distribution f_Π.

Mathematically, a parthood distribution is a Boolean function from the powerset of $\{1, \ldots, n\}$ to $\{0, 1\}$, i.e. it takes a collection of source indices as an input and returns either 0 (the atom described by the distribution is not part of information provided by the collection) or 1 (the atom described by the distribution is part of that information) as an output. But note that not all such functions qualify as a parthood distribution. We already saw that certain constraints have to be satisfied. For instance, the empty set of sources has to be mapped to 0. We propose that there are exactly three constraints a parthood distribution f has to satisfy leading to the following definition:

Definition 2.1. A parthood distribution is any function $f : \mathcal{P}(\{1, \ldots, n\}) \to \{0, 1\}$ such that

(i) $f(\{\}) = 0$ (There is no information in the empty set)
(ii) $f(\{1, \ldots, n\}) = 1$ (All information is in the full set)
(iii) For any two collections of source indices $\mathbf{a}$, $\mathbf{b}$: If $\mathbf{b} \supseteq \mathbf{a}$, then $f(\mathbf{a}) = 1 \Rightarrow f(\mathbf{b}) = 1$ (Monotonicity).

The third constraint says that if an atom of information is part of the information provided by some collection of sources $\mathbf{a}$, then it also has to be part of the information provided by any superset of this collection. For example, if an atom is part of the information in source 1, then it also has to be part of the information in sources 1 and 2 combined. Note that this monotonicity constraint only matters if there are more than two information sources. Otherwise it is implied by the first two constraints. To fix ideas, an example of a Boolean function that is *not* a parthood distribution is shown in table 2. The function assigns a 1 to the collection $\{1\}$ but a 0 to collections $\{1, 2\}$ and $\{1, 3\}$ which are supercollections of $\{1\}$. Thus, there can be no atom of information with the parthood relations described by this Boolean function.

We may now answer the question about the meaning of the atoms of information, i.e. what *type* of information they represent: they represent information that is part of the information provided by certain collections of sources but not part of the information of other collections. More precisely, we can phrase this idea in terms of the following core principle:

Principle 2.2. Each atom of information is characterized by a parthood distribution describing whether or not it is part of the information provided by the different possible collections of sources. The atom $\Pi(f)$ with parthood distribution f is exactly that part of the joint mutual information about the target which is (1) part of the information provided by all collections of sources $\mathbf{a}$ for which $f(\mathbf{a}) = 1$, and (2), which is not part of the information provided by collections for which $f(\mathbf{a}) = 0$.

Given this characterization of the information atoms we are now in a position to answer the second question: how many atoms are there for a given number of information sources.

Table 3. The two constant Boolean functions are ruled out by the first and second constraint on parthood distributions described above.

part of	{}	. . .	. . .	. . .	{1, . . . , *n*}
	1	**1**	**1**	**1**	**1**
	0	**0**	**0**	**0**	0

(b) How many atoms of information are there?

Since each atom is characterized by its parthood distribution, the answer is straightforward: there is one atom per parthood distribution, or in other words, one atom per Boolean function satisfying the constraints presented in the previous section. The monotonicity constraint turns out to be most restrictive. In fact, once the monotonicity constraint is satisfied the other two constraints only rule out one Boolean function each as shown in table 3. The reason is the following: Firstly, there is only a single *monotonic* Boolean function that assigns the value 1 to the empty set, namely, the function that is always 1. Since the empty set is subset of any other set, monotonicity enforces to assign a 1 to all sets once the empty set has value 1. However, this possibility is ruled out by the first constraint saying that there is no information in the empty set. Secondly, there is only a single *monotonic* Boolean function assigning the value 0 to the full set $\{1, \ldots, n\}$, namely the function that is always 0. Since any other set of source indices is contained in the full set, monotonicity forces us to assign a 0 to all sets once the full set has value 0. If we were to assign a 1 to any other set, then we would have to assign a 1 to the full set as well.

This means that the number of atoms is equal to *the number of monotonic Boolean functions minus two*. Now the sequence of the numbers of monotonic Boolean functions of n-bits is a very famous sequence in combinatorics called the *Dedekind numbers*. The Dedekind numbers are a very rapidly (in fact super-exponentially) growing sequence of numbers of which only the first eight entries are known to date [17]. The values for $2 \leq n \leq 6$ of the Dedekind numbers are: 6, 20, 168, 7581, 7828354.

Now that we have answered what type of information the different atoms represent and how many there are for a given number of information sources, there is one important question left: How large are these different atoms? How many *bits* of information does each atom provide?

(c) How large are the atoms of information?

The question of the sizes of the atoms is not a trivial one since the number of atoms grows so quickly. In the case of four information sources, there are already 166 atoms. Hence, it does not appear to be feasible to define the amount of information of each of these atoms separately. What we need is a systematic approach that somehow fixes the sizes of all atoms at the same time. The core idea is to transform the problem into a much simpler one in which only a single type of informational quantity has to be defined. In the following, we show how this can be achieved in three steps.

(i) Define a quantitative relationship between atoms and composite quantities

So far we have only discussed how the atoms of information relate *qualitatively* to composite information quantities that are made up of multiple atoms, in particular mutual information (in the next section we will encounter another non-atomic quantity). We saw for instance, that in the case of two sources, the mutual information contributions provided by the individual sources, $I(T : S_1)$ and $I(T : S_2)$, each consist of a unique and a redundant information atom, while the joint mutual information $I(T : S_1, S_2)$ additionally consists of a synergistic part. This is illustrated in the information diagram shown in figure 2.

Now the question arises: How are these mutual information terms related to the atoms they consist of *quantitatively*? The most straightforward answer (and the one generally accepted in

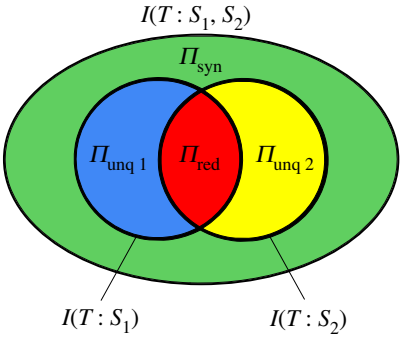

Figure 2. Information diagram depicting the partial information decomposition for the case of two information sources. The inner two black circles represent the mutual information provided by the first source (left) and the second source (right) about the target. Each of these mutual information terms contains two atomic parts: $I(T : S_1)$ consists of the unique information in source 1 ($\Pi_{\text{unq 1}}$, blue patch) and the information shared with source 2 (Π_{red}, red patch). $I(T : S_2)$ consists of the unique information in source 2 ($\Pi_{\text{unq 2}}$, yellow patch) and again the shared information. The joint mutual information $I(T : S_1, S_2)$ is depicted by the large black oval encompassing the inner two circles. $I(T : S_1, S_2)$ consists of four atoms: the unique information in source 1 ($\Pi_{\text{unq 1}}$, blue patch), the unique information in source 2 ($\Pi_{\text{unq 2}}$, yellow patch), the shared information (Π_{red}, red patch) and additionally the synergistic information (Π_{syn}, green patch). (Online version in colour.)

the PID field) is that the mutual information is simply the *sum* of the atoms it consists of. We propose to extend this principle to any composite information quantity, i.e. any quantity that can be described as being made up out of multiple information atoms.

Principle 2.3. The size of any non-atomic information quantity (i.e. the amount of information it contains) is the sum of the sizes of the information atoms it consists of.

We could also rephrase this as 'wholes are the sums of their (atomic) parts'. In the case of two information sources, this principle leads to the following three equations:

$$I(T : S_1, S_2) = \Pi_{\text{red}} + \Pi_{\text{unq 1}} + \Pi_{\text{unq 2}} + \Pi_{\text{syn}} \tag{2.2}$$

$$I(T : S_1) = \Pi_{\text{red}} + \Pi_{\text{unq 1}} \tag{2.3}$$

$$I(T : S_2) = \Pi_{\text{red}} + \Pi_{\text{unq 2}}. \tag{2.4}$$

This already gets us quite far in terms of determining the sizes of the atoms: the sizes of the atoms are the solutions to a linear system of equations. The only problem is that the system is underdetermined. We have four unknowns but only three equations. In the case of three sources, the problem is even more severe. In this case, there are seven non-empty collections of sources, and, hence, seven mutual information terms. Again each of these terms is the sum of certain atoms. But as shown in section §2b there are 18 atoms. So we are short of 11 equations!

In general, the equations relating the mutual information provided by some collection of sources $\mathbf{a}$ and the information atoms can be expressed easily in terms of their parthood distributions

$$I(T : \mathbf{a}) = \sum_{f(\mathbf{a})=1} \Pi(f), \tag{2.5}$$

where $\Pi(f)$ is the information atom corresponding to parthood distribution f and the summation notation means that we are summing over all f such that $f(\mathbf{a}) = 1$. Note that on the left-hand side, we are using the shorthand notation $I(T : \mathbf{a})$ for the mutual information $I(T : (S_i)_{i\in\mathbf{a}})$ provided by the collection $\mathbf{a}$. Equation (2.5) can be taken to define a minimal notion of a PID, i.e. any set of quantities $\Pi(f)$ at least has to satisfy this equation in order to be considered a PID (or at least to be considered a parthood-based/Williams and Beer type PID). For a formal definition of such a minimally consistent PID see electronic supplementary material, appendix §1.

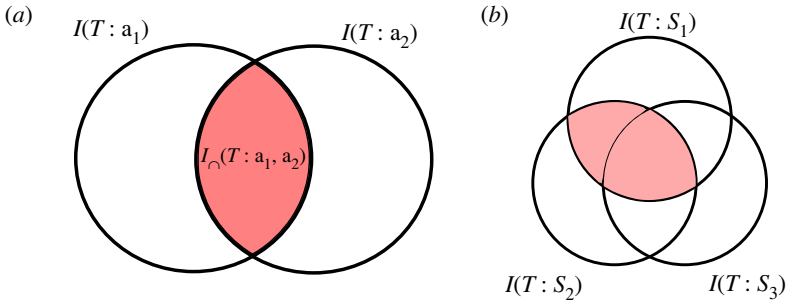

Figure 3. (*a*) Illustration of the idea of the redundant information of collections a_1 and a_2. (*b*) Redundant information is generally not an atomic quantity. In the context of three information sources, the redundant information of sources 1 and 2 consists of two parts: the information shared by *only* by sources 1 and 2, and the information shared by all three sources. (Online version in colour.)

This concludes the first step. The next one is to find a way to come up with the appropriate number of additional equations. In doing so we will follow the same approach as Williams and Beer and use the concept of *redundant information* to introduce additional constraints. It should be noted that this is not the only way to derive a solution for the information atoms. In other words, a PID does not have to be 'redundancy based'. This issue is discussed in detail in §5. For now, however, let us follow the conventional path and see how it enables us to determine the sizes of the atoms of information.

(ii) Formulate additional equations using the concept of redundant information

The basic idea is now to extent the considerations of the previous step to another composite information quantity: the redundant information provided by multiple collections of sources about the target which we will generically denote by $I_\cap(T : \mathbf{a}_1, \ldots, \mathbf{a}_m)$. The $\cap$ symbol refers to the idea that the redundant information of collections $\mathbf{a}_1, \ldots, \mathbf{a}_m$ is the information contained in $\mathbf{a}_1$ *and* $\mathbf{a}_2$ *and*, …, *and* $\mathbf{a}_m$. Intuitively, given two collections of sources $\mathbf{a}_1$ and $\mathbf{a}_2$, their redundant information is the information 'shared' by those collections, what they have 'in common', or geometrically: their overlap. These informal ideas are illustrated on the left side in figure 3.

Note that the redundant information of multiple collections of information sources is not defined in classical information theory. We have to come up with an appropriate measure of redundant information ourselves. However, the informal ideas just describes already tell us that redundant information, no matter how we define it, should be related qualitatively to the information atoms in a very specific way: the information redundantly provided by multiple collections of sources should consist of exactly those information atoms that are part of the information carried by *all* of those collections:

Principle 2.4. The redundant information $I_\cap(T : \mathbf{a}_1, \ldots, \mathbf{a}_m)$ consists of all information atoms that are part of the information provided by *each* $\mathbf{a}_i$, i.e. all atoms with a parthood distribution satisfying $f(\mathbf{a}_i) = 1$ for all $i = 1, \ldots, m$.

Let us see what this principle implies in concrete examples. We saw that in the case of two sources, the redundant information of sources 1 and 2, $I(T : \{1\}, \{2\})$, is actually itself an atom, namely the atom with the parthood distribution

{}	{1}	{2}	{1, 2}
0	1	1	1

This is the only atom that is part of both the information provided by the first source and also part of the information provided by the second source. But this is really a special case. Note what happens if we add a third source to the scenario. In this case, the redundant information

$I(T:\{1\},\{2\})$ of sources 1 and 2 should consist of *two* parts: first, the information shared by *all three* sources (which is certainly also shared by sources 1 and 2), and, second, the information shared *only* by sources 1 and 2 but not by source 3. This is illustrated on the right side in figure 3. Note also that in the case of three sources there are actually *many* redundancies that we may compute:

(i) the redundancy of all three sources $I_\cap(T:\{1\},\{2\},\{3\})$;
(ii) the redundancy of any *pair* of sources such as the redundancy of $I_\cap(T:\{1\},\{2\})$;
(iii) the redundancy between a single source and a pair of sources such as $I_\cap(T:\{1\},\{2,3\})$;
(iv) the redundancy between two pairs of sources such as $I_\cap(T:\{1,2\},\{2,3\})$;
(v) the redundancy of all three possible pairs of sources $I_\cap(T:\{1,2\},\{1,3\},\{2,3\})$.

It turns out that in total there are 11 redundancies (strictly speaking we should say 11 'proper' redundancies as will be explained below). But this is exactly the number of missing equations in the case of three information sources (see last paragraph of previous section).

Now, combining principles 2.3 and 2.4, allows us the answer what the *quantitative* relationship between redundant information and information atoms has to be: the redundant information of collections of sources $\mathbf{a}_1,\ldots,\mathbf{a}_m$ is the sum of all atoms that are part of the information provided by *each* collection:

$$I_\cap(T:\mathbf{a}_1,\ldots,\mathbf{a}_m)=\sum_{f(\mathbf{a}_i)=1\forall i=1,\ldots,m}\Pi(f), \tag{2.6}$$

where again the notation means that we are summing over all f that satisfy the condition below the summation sign. This equation can be read in two ways: first, as placing a constraint on the redundant information $I_\cap$, namely that it has to be the sum of specific atoms. This means that if we already knew the sizes of the Π's, we could compute $I_\cap$. However, the sizes of the Π's are precisely what we are trying to work out. Now the crucial idea is that we can also read the equation the other way around: if we can come up with some reasonable measure of redundant information $I_\cap$ we may be able to *invert* equation (2.6) in order to obtain the Π's. So the final step will be to show that such an inversion is in fact possible and will lead to a unique solution for the atoms of information.

Before proceeding to this step, it is important to briefly clarify the relationships between the three central concepts we have discussed so far:

(i) the mutual information (the quantity we want to decompose)
(ii) the information atoms (the quantities we are looking for)
(iii) redundant information (the quantity we are going to use to find the information atoms).

These concept are easily confused with each other but should be clearly separated. The relationships between them are shown in figure 4. First, based on what we have said so far, mutual information can be shown to be a special case of redundant information: the redundant information of a single collection $I_\cap(T:\mathbf{a}_1)$, i.e. 'the information the collection shares *with itself* about the target'. The reason for this is that principle 2.4 tells us that the redundant information of a single collection consists of all the atoms that are part of the mutual information carried by *that* collection about the target. But this is simply the mutual information of that collection

$$I_\cap(T:\mathbf{a}_1)\overset{\text{Eq.(2.6)}}{=}\sum_{f(\mathbf{a}_i)=1\forall i=1,\ldots,m}\Pi(f)=\sum_{f(\mathbf{a}_1)=1}\Pi(f)\overset{\text{Eq.(2.5)}}{=}I(T:\mathbf{a}_1). \tag{2.7}$$

Accordingly, mutual information has been called 'self-redundancy' in the PID literature (although not based on parthood arguments) [13]. The relationship between redundant information and atoms is as follows: only the 'all-way' redundancy, i.e. the information shared by *all* n sources is itself an atom. Any other redundancy, such as the redundancy of only a subset of sources, is a composite quantity made up out of multiple atoms.

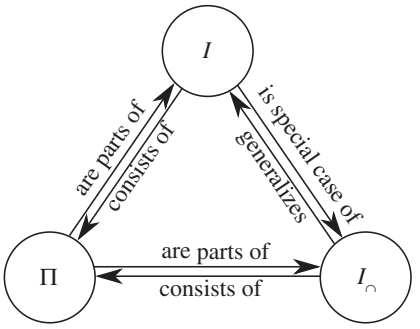

Figure 4. Relationships between mutual information, redundant information and information atoms. (Online version in colour.)

(iii) Show that a measure of redundant information leads to a unique solution for the information atoms

There is a very useful fact about parthood distributions that will help us to obtain a unique solution for the atoms given an appropriate measure of redundant information: parthood distributions can be ordered in a very natural way into a lattice structure that is tightly linked to the idea of redundancy. The lattice for the case of three sources is shown in figure 5. The parthood distributions are ordered as follows: if there is a 1 in certain positions on a parthood distribution f, then all the parthood distributions g below it also have a 1 in the same positions, plus some additional ones. Or in terms of the atoms corresponding to these parthood distributions: If an atom $\Pi(f)$ is part of the information provided by some collections of sources, then all the atoms $\Pi(g)$ below it are also part of the information provided by these collections. Formally, we will denote this ordering by $\sqsubseteq$ and it is defined as

$$f \sqsubseteq g \Leftrightarrow (f(\mathbf{a}) = 1 \rightarrow g(\mathbf{a}) = 1 \text{ for any } \mathbf{a} \subseteq \{1, \ldots, n\}). \tag{2.8}$$

For n information sources, we will denote the lattice of parthood distributions by $(\mathcal{B}_n, \sqsubseteq)$, where $\mathcal{B}_n$ is the set of all parthood distribution in the context of n sources (for proof that this structure is in fact a lattice in the formal sense see electronic supplementary material, appendix §2).

Note that the different 'levels' of the lattice contain parthood distributions with the same number of ones and that higher level parthood distributions contain *less* ones: At the very top in figure 5, there is the parthood distribution describing the atom that is *only* part of the joint mutual information provided by all three sources combined, i.e. the synergy of the three sources. One level down, there are the three parthood distributions that assign the value 1 exactly two times. Yet another level down, we find the three possible parthood distributions that assign the value 1 exactly three times. And so on and so forth until we reach the bottom of the lattice which corresponds to the information shared by all three sources. Accordingly, the corresponding parthood distribution assigns the value 1 to all collections (except of course the empty collection).

Ordering all the parthood distributions (and hence atoms) into such a lattice provides a good overview that tells us how many atoms exist for a given number of source variables and what their characteristic parthood relationships are. But the lattice plays a much more profound role because it is very closely connected to the concept of redundant information. The idea is to associate with each parthood distribution in the lattice a particular redundancy: the redundant information of all the collections that are assigned the value 1 by the distribution. In other words, for any parthood distribution f we consider the redundancy

$$I_\cap(T : f) := I_\cap\left(T : (\mathbf{a} \mid f(\mathbf{a}) = 1)\right). \tag{2.9}$$

For example, in the case of three sources, the redundant information associated with the parthood distribution that assigns value 1 to collections $\{1,2\}$, $\{2,3\}$ and $\{1,2,3\}$, and value 0 to all other collections (the one emphasized in figure 5), is simply $I_\cap(T : \{1,2\}, \{2,3\}, \{1,2,3\})$. We saw in the

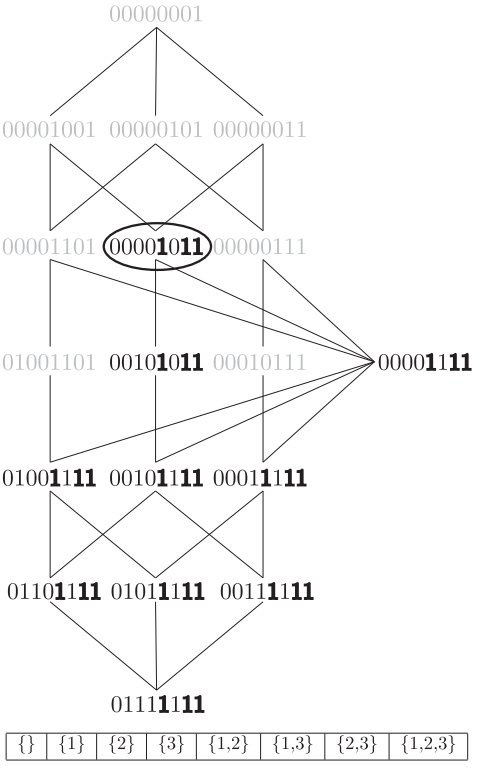

Figure 5. Lattice of parthood distributions for the case of three information sources. The parthood distributions are represented as bit-strings where the *i*-th bit is the value that the parthood distribution assigns to the *i*-th collection of sources. The order of these collections is shown below the lattice for reference. A distribution *f* is below a distribution *g* just in case *f* has value 1 in the same positions as g and in some additional positions. This is illustrated for the parthood distribution highlighted by the black circle. The positions in which it assigns the value 1 are marked in bold.

previous section that any redundancy $I_\cap(T : \mathbf{a}_1, \ldots, \mathbf{a}_m)$ is the sum of all atoms that are part of the information provided by each of the $\mathbf{a}_i$. Now here is the connection between the lattice and redundant information: these atoms are the ones that have value 1 on each $\mathbf{a}_i$. But, by definition of the ordering, these are precisely the ones corresponding to parthood distributions *below and including* the parthood distribution for which we are computing the associated redundancy. In other words, the redundant information associated with a parthood distribution f can always be expressed as

$$I_\cap(T : f) = \sum_{g \sqsubseteq f} \Pi(g). \tag{2.10}$$

In this way, we obtain one equation per parthood distribution. And since there are as many information atoms as parthood distributions, we obtain as many equations as unknowns. This is already a good sign. But is a unique solution for the information atoms guaranteed? This question can be answered affirmatively by noting that the system of equations described by (2.10) (one equation per f) is not just any linear system, but has a very special structure: one function $I_\cap(T : f)$ evaluated at a point f on a lattice is the sum of another function $\Pi(f)$ over all points on the lattice below and including the point f. The process of solving such a system for the $\Pi(f)$'s once all the $I_\cap(T : f)$'s are given, or in other words *inverting* equation (2.10), is called *Moebius inversion*. Crucially, a unique solution is guaranteed for any real or even complex valued function $I_\cap$ that we may put on the lattice [18].

This means that we have now completely shifted the problem of determining the sizes of the information atoms to the problem of coming up with a reasonable definition of redundant

information $I_\cap(T:f)$. Even though we have to define this quantity for *each* parthood distribution f this is still a much simpler task. The reason is that all the $I_\cap$'s represent exactly the same *type* of information, namely redundant information. On the other hand, the information atoms Π represent completely different types of information. Even in the simplest case of two sources we have to deal not only with redundant information, but also unique information and synergistic information. And the story gets more and more complicated the more information sources are considered.

Now, note that apparently we only need to define quite special redundant information terms, namely the redundancies associated with parthood distributions $I_\cap(T:f)$ (see definition (2.9)). However, we will now show that these are in fact *all* possible redundancies, i.e. the redundancy of any tuple of collections of sources $\mathbf{a}_1,\dots,\mathbf{a}_m$ is necessarily equal to a redundancy associated with a specific parthood distribution. The reason for this is that the quantitative relation between atoms and redundant information (equation (2.6)) not only provides a way to solve for the information atoms once we know $I_\cap$, it also implies that $I_\cap$ has to satisfy the following invariance properties:

(i) $I_\cap(T:\mathbf{a}_1,\dots,\mathbf{a}_m) = I_\cap(T:\mathbf{a}_{\sigma(1)},\dots,\mathbf{a}_{\sigma(m)})$ for any permutation σ (**symmetry**).
(ii) If $\mathbf{a}_i = \mathbf{a}_j$ for $i \neq j$, then $I_\cap(T:\mathbf{a}_1,\dots,\mathbf{a}_m) = I_\cap(T:\mathbf{a}_1,\dots,\mathbf{a}_{i-1},\mathbf{a}_{i+1},\dots,\mathbf{a}_m)$ **(idempotency)**.
(iii) If $\mathbf{a}_i \supset \mathbf{a}_j$ for $i \neq j$, then $I_\cap(T:\mathbf{a}_1,\dots,\mathbf{a}_m) = I_\cap(T:\mathbf{a}_1,\dots,\mathbf{a}_{i-1},\mathbf{a}_{i+1},\dots,\mathbf{a}_m)$ **(invariance under superset removal / addition)**.
(iv) $I_\cap(T:\mathbf{a}) = I(T:\mathbf{a})$ (**self-redundancy**).

We can easily ascertain that any measure of redundant information $I_\cap$ has to have these properties by taking a closer look at the condition describing which atoms to sum over in order to obtain a particular redundant information term $I(T:\mathbf{a}_1,\dots,\mathbf{a}_m)$: we have to sum over the atoms with parthood distribution satisfying $f(\mathbf{a}_i)=1$ for all $i=1,\dots,m$. Now whether or not this condition is true of a given parthood distribution f, first, does not depend on the *order* in which the collections $\mathbf{a}_i$ are given (symmetry), secondly, it does not depend on whether the same collection $\mathbf{a}$ is repeated multiple times (idempotency), and thirdly, it does not matter whether we add or remove some collection $\mathbf{a}_i$ that is a proper superset of some other collection (superset removal/addition). This fact is due to the monotonicity constraint on parthood distributions. Finally, the 'self-redundancy' property was already established in the previous section.

These invariance properties are referred in the literature as the Williams and Beer axioms for redundant information [14] (in addition there is a *quantitative* monotonicity axiom that we reject. See §6). However, in the parthood formalism described here they are not themselves axioms but are *implied* by the core principles we have set out. The first two invariance properties imply that we may restrict ourselves to *sets* instead of tuples of collections in defining $I_\cap$. The third constraint additionally tells us that we can restrict ourselves to those sets of collections $\{\mathbf{a}_1,\dots,\mathbf{a}_m\}$ such that no collection $\mathbf{a}_i$ is a superset of another collection $\mathbf{a}_j$. Such sets of collections are called *antichains*. Hence, the redundancy of *any* tuple of collections of sources $\mathbf{a}_1,\dots,\mathbf{a}_m$ is necessarily equal to the redundancy associated with a particular antichain. This antichain results from ignoring the order and repetitions of the $\mathbf{a}_i$, and removing any supersets. For instance, $I_\cap(T:\{1\},\{1\},\{2\},\{1,2\}) = I_\cap(T:\{1\},\{2\})$.

We can now see that the redundancies $I_\cap(T:f)$ are in fact all possible redundancies by associating with any antichain $\alpha = \{\mathbf{a}_1,\dots,\mathbf{a}_m\}$ a parthood distribution f_α that assigns the value 1 to all $\mathbf{a}_i$ *and all supersets of these collections*, while it assigns the value 0 to all other collections. Now, due to the invariance of $I_\cap$ under removal of supersets, it immediately follows that $I_\cap(T:f_\alpha) = I_\cap(T:\alpha)$. So in conclusion, there is one redundancy for each antichain α and these redundancies are equal to the redundancies associated with the corresponding parthood distributions. Hence the redundancies $I_\cap(T:f)$ are in fact *all* possible redundancies.

Of course, there is also an inverse mapping associating with any parthood distribution f an antichain α_f. In fact, the lattice of parthood distributions $(\mathcal{B}_n,\sqsubseteq)$ is *isomorphic* to a lattice of antichains $(\mathcal{A}_n,\preceq)$ equipped with an ordering relationship that was originally introduced by Crampton & Loizou [19] and used by Williams & Beer in their original exposition of PID.

The formal proof of this fact is postponed to §4 where a third perspective on PID, the logical perspective, is introduced.

In the next section, we will tackle the problem of defining a measure of redundant information for each parthood distribution/antichain by connecting PID theory to formal logic. The measure $I_\cap^{\mathrm{sx}}$ derived in this way is identical to the one described in [20]. In showing how this measure can be inferred from logical and parthood principles we aim to (i) strengthen the argument for $I_\cap^{\mathrm{sx}}$, and (ii) open the gateway between PID theory and formal logical. This latter point is elaborated in §4.

3. Using logic to derive a measure of redundant information

We have now solved the PID problem up to specifying a reasonable measure of redundant information $I_\cap$ between collections that form an antichain. In this section, we will derive such a measure. In doing so we will first move from the level of random variables $T, S_1, \ldots, S_n$ to the level of particular realizations $t, s_1, \ldots, s_n$ of these variables. This level of description is generally called the *pointwise* level and has been used as the basis of classical information theory by Fano [21]. Pointwise approaches to PID have been put forth by [14,20].

Note that moving to the level of realizations simplifies the problem considerably because realizations are much simpler objects than random variables. A realization is simply a specific symbol or number whereas a random variables is an object that may take on various different values and can only be fully described by an entire probability distribution over these values.

(a) Going pointwise

The idea underlying the pointwise approach is to consider the information provided by a particular joint realization (observation) of the source random variables about a particular realization (observation) of the target random variable. So from now on we assume that these variables have taken on *specific* values $s_1, \ldots, s_n, t$. It was shown by Fano [21] that the whole of classical information theory can be derived from this pointwise level. By placing a certain number of reasonable constraints or axioms on pointwise information, it follows that this information must have a specific form. In particular, the pointwise mutual information $i(t:s)$ is given by

$$i(t:s) := \log\left(\frac{P(t|s)}{P(t)}\right). \tag{3.1}$$

The mutual information $I(T:S)$ is then simply defined as the *average* pointwise mutual information. Note that pointwise mutual information (in contrast to mutual information) can be both positive and negative. It essentially measures whether we are guided in the right or wrong direction with the respect to the actual target realization t. If the conditional probability of $T=t$ given the observation of $S=s$ is larger than the unconditional (prior) probability of $T=t$, then we are guided in the right direction: the actual target realization is in fact t and observing that $S=s$ makes us more likely to think so. Accordingly, in this case, the pointwise mutual information is *positive*. On the other hand, if the conditional probability of $T=t$ given the observation of $S=s$ is smaller than the unconditional (prior) probability of $T=t$, then we are guided in the wrong direction: observing $S=s$ makes us less likely to guess the correct target value. In this case, the pointwise mutual information is *negative*. The joint pointwise mutual information of source realizations $s_1, \ldots, s_n$ about the target realization is defined in just the same way

$$i(t:s_1,\ldots,s_n) := \log\left(\frac{P(t|s_1,\ldots,s_n)}{P(t)}\right). \tag{3.2}$$

The idea is now to perform the entire PID on the pointwise level, i.e. to decompose the pointwise joint mutual information $i(t:s_1,\ldots,s_n)$ that the source realizations provide about the target realization [14]. This leads to *pointwise atoms* $\pi_{s_1,\ldots,s_n,t}$ (in the following we will generally drop the subscript). Crucially, we are only changing the quantity to be decomposed from

$I(T : S_1, \ldots, S_n)$ to $i(t : s_1, \ldots, s_n)$. Otherwise, the idea is completely analogous to what we have discussed in §2 (simply replace I by i and Π by π): the goal is to decompose the pointwise mutual information into information atoms that are characterized by their parthood relations to the pointwise mutual information provided by the different possible collections of source realizations. These atoms have to stand in the appropriate relationship to *pointwise redundancy*: the pointwise redundancy $i_\cap(t : \mathbf{a}_1, \ldots, \mathbf{a}_m)$ is the sum of all pointwise atoms $\pi(f)$ that are part of the information provided by *each* collection of source realizations $\mathbf{a}_i$. By exactly the same argument as described in §2c(iii), there is a unique solution for the pointwise atoms once a measure of pointwise redundancy $i(t : \alpha)$ is fixed for all antichains $\alpha = \{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$. The *variable-level* atoms Π are then defined as the *average* of the corresponding pointwise atoms

$$\Pi(f) = \sum_{s_1,\ldots,s_n,t} P(s_1, \ldots, s_n, t)\pi_{s_1,\ldots,s_n}(f). \tag{3.3}$$

We are now left with defining the pointwise redundancy $i_\cap$ among collections of source realizations. As noted above this is a much easier problem than coming up with a measure of redundancy among collections of entire source variables. In the next section, we show how the pointwise redundancy of multiple collections of source realizations can be expressed as the information provided by a particular *logical statement* about these realizations.

(b) Defining pointwise redundancy in terms of logical statements

The language of formal logic allows us to form statements about the source realizations. In particular, we will consider statements made up out of the following ingredients:

(i) n basic statements of the form $S_i = s_i$, i.e. 'Source S_i has taken on value s_i'
(ii) the logical connectives $\wedge$ (and), $\vee$ (or), $\neg$ (not), $\rightarrow$ (if, then)
(iii) brackets),(

In this way, we may form statements such as $S_1 = s_1 \wedge S_2 = s_2$ (source S_1 has taken on value s_1 and source S_2 has taken on value s_2) or $S_1 = s_1 \vee (S_2 = s_2 \wedge S_3 = s_3)$ (either source S_1 has taken on value s_1 or source S_2 has taken on value s_2 and source S_3 has taken on value s_3). Now we may ask: What is the information provided by the truth of such statements about the target realization t? Classical information theory allows us to quantify this information as a pointwise mutual information: Let A be any statement of the form just described, then the information $i(t : A)$ provided by the truth of this statement is

$$i(t : A) := i(t : \mathbb{I}_A = 1) = \log\left(\frac{P(t|A \text{ is true})}{P(t)}\right), \tag{3.4}$$

where $\mathbb{I}_A$ is the *indicator random variable* of the event that the statement A is true, i.e. $\mathbb{I}_A = 1$ if the event occurred and $\mathbb{I}_A = 0$ if it did not. The interpretation of this information is that it measures whether and to what degree we are guided in the right or wrong direction with respect to the actual target value once we learn that statement A is true.

Note that according to this definition the pointwise mutual information provided by a collection of source realizations $i(t : \mathbf{a})$ is the information provided by the truth of the *conjunction* $\bigwedge_{i\in\mathbf{a}} S_i = s_i$:

$$i(t : \mathbf{a}) = i\left(t : \bigwedge_{i\in\mathbf{a}} S_i = s_i\right). \tag{3.5}$$

Therefore, the information redundantly provided by collections of source realizations $\mathbf{a}_1, \ldots, \mathbf{a}_m$ is the information redundantly provided by the truth of the corresponding conjunctions. Now, what is this information? We propose that in general the following principle describes redundancy among statements:

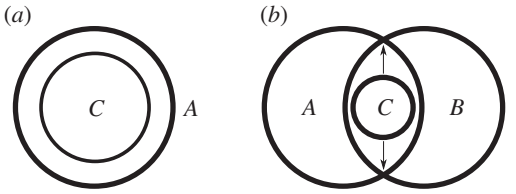

Figure 6. (*a*) Information diagram depicting the information provided by statement *A* and *C*. If statement *C* is logically weaker than statement *A*, i.e. if *C* is implied by *A*, then the information provided by *C* has to be part of the information provided by *A*. (*b*) Information diagram depicting the information provided by statements *A*, *B* and *C*. *C* is assumed to be logically weaker than both *A* and *B*. Thus it has to be part of the information provided by *A* and also part of the information provided by *B*. Accordingly, it is contained in the 'overlap', i.e. the redundant information of *A* and *B*. In order to obtain the entire redundant information statement *C* has to be 'maximized', i.e. it has to be chosen as the strongest statement weaker than both *A* and *B* (this is indicated by the arrows).

Principle 3.1. The information redundantly provided by the truth of the statements $A_1, \ldots, A_m$ is the information provided by the truth of their *disjunction* $A_1 \vee \ldots \vee A_m$.

There are two motivations for this principle. First, the logical inferences to be drawn from the disjunction $A \vee B$ are precisely the inferences that can be drawn *redundantly* from both A and B. If some conclusion C logically follows from both A and B, then it also follows from $A \vee B$. Conversely, any conclusion C that follows from the disjunction $A \vee B$ follows from both A and B. Formally,

$$A \vee B \models C \Leftrightarrow A \models C \text{ and } B \models C, \tag{3.6}$$

where $\models$ denotes logical implication. The second motivation again invokes the idea of parthood relationships: *If some statement C is logically weaker than a statement A, then the information provided by C should be part of the information provided by A*. For instance, the information provided by the statement $S_1 = s_1$ has to be part of the information provided by the statement $S_1 = s_1 \wedge S_2 = s_2$. This idea is illustrated in the information diagram on the left side in figure 6.

Now, this idea implies that if a statement C is weaker than *both* A and B, then the information provided by C is part of the information carried by A and also part of the information carried by B. But this means that the information provided by C is part of the *redundant information* of A and B. In order to obtain the *entire* redundant information, the statement C should therefore be chosen as the *strongest* statement logically weaker than both A and B (see right side of figure 6). But this statement is the disjunction $A \vee B$ (or any equivalent statement).

Based on these ideas, we can now finally formulate our proposal for a measure of pointwise redundancy $i_\cap(t : \mathbf{a}_1, \ldots, \mathbf{a}_m)$. We noted above that the information redundantly provided by collections of realizations $\mathbf{a}_1, \ldots, \mathbf{a}_m$ is the information redundantly provided by the conjunctions $\bigwedge_{i \in \mathbf{a}_j} S_i = s_i$. And by the arguments just presented this is the information provided by the *disjunction of these conjunctions*. We denote the function that measures pointwise redundant information in this way by $i_\cap^{\mathrm{sx}}$ (for reasons that will be explained shortly). It is formally defined as

$$i_\cap^{\mathrm{sx}}(t : \mathbf{a}_1, \ldots, \mathbf{a}_m) := i\left(t : \bigvee_{j=1}^{m} \bigwedge_{i \in \mathbf{a}_j} S_i = s_i\right). \tag{3.7}$$

Recall that by definition this is the pointwise mutual information provided by the truth of the statement in question. Hence, it measures whether and to what degree we are guided in the right or wrong direction with respect to the actual target value t once we learn that the statement is true.

We have now arrived at a *complete* solution to the PID problem: given the measure $i_\cap^{\mathrm{sx}}$ we may carry out the Moebius inversion

$$i_\cap^{\mathrm{sx}}(t : f) = \sum_{g \sqsubseteq f} \pi^{\mathrm{sx}}(f), \tag{3.8}$$

in order to obtain the pointwise atoms π^{sx}. This has to be done for *each* realization $s_1, \ldots, s_n, t$. The obtained values can then be averaged as per equation (3.3) to obtain the variable-level atoms Π^{sx}.

As shown in [20], the measure $i_\cap^{\mathrm{sx}}$ can also be motivated in terms of the notion of *shared exclusions* (hence the superscript 'sx'). The underlying idea is that redundant information is linked to possibilities (i.e. points in sample space) that are redundantly excluded by multiple source realizations. We argue that the fact that the measure $i_\cap^{\mathrm{sx}}$ can be derived in these two independent ways provides further support for its validity. We offer a freely accessible implementation of the $i_\cap^{sx}$ PID as part of the IDTxl toolbox [22]. Worked examples of its computation and details on the computational complexity can be found in [20].

In the following section, we show that the value of formal logic within the theory of PID goes far beyond helping us to define a measure of pointwise redundant information. In fact, similar to the lattices of parthood distributions and antichains, there is a lattice of logical statements that can equally be used as the basic mathematical structure of PID. This lattice is particularly useful because the ordering relationship turns out to be very simple and well understood: the relation of logical implication. We will show that this perspective also offers an independent starting point for the development of PID theory.

4. The logical perspective

(a) Logic lattices

The considerations of the previous section identified the information redundantly provided by collections $\mathbf{a}_1, \ldots, \mathbf{a}_m$ with the information provided by a particular logical statement: a disjunction of conjunctions of basic statements of the form $S_i = s_i$. This has an interesting implication: there is a one-to-one mapping between antichains α and logical statements. Let us now look at this situation a bit more abstractly by replacing the concrete statements $S_i = s_i$ with *propositional variables* $\phi_1, \ldots, \phi_n$. Together with the logical connectives $\neg, \vee, \wedge, \rightarrow$ (plus brackets) these form a language of propositional logic [23]. We will denote this language by $\mathbb{L}$. We may now formally introduce a mapping Ψ from the set of antichains $\mathcal{A}$ into $\mathbb{L}$ via

$$\Psi : \mathcal{A} \to \mathbb{L}, \quad \text{where } \alpha \rightarrowtail \tilde{\alpha} := \bigvee_{a \in \alpha} \bigwedge_{i \in a} \phi_i. \tag{4.1}$$

In other words, α is mapped to a statement by first conjoining the ϕ_i corresponding to indices *within* each $\mathbf{a}_i$ and then disjoining these conjunctions. For instance, the antichain $\{\{1,2\},\{2,3\}\}$ will be associated with the statement $(\phi_1 \wedge \phi_2) \vee (\phi_2 \wedge \phi_3)$. Note that if we interpret the propositional variables ϕ_i as 'source S_i has taken on value s_i', then this is of course precisely the mapping of an antichain to the statement providing the redundant information (in the sense of $i_\cap^{sx}$) associated with that antichain.[1]

The range $\mathcal{L} \subseteq \mathbb{L}$ of Ψ is *set of all disjunctions of logically independent conjunctions of pairwise distinct propositional variables*. The logical independence of the conjunctions is the logical counterpart of the antichain property. The 'pairwise distinct' condition ensures that the same atomic statement does not occur multiple times in any conjunction. The set $\mathcal{L}$ can now be equipped with the relationship of logical implication $\models$ in order to obtain a new structure $(\mathcal{L}, \dashv)$ which we will show to be isomorphic to the lattices of antichains and parthood distributions. Here $\models$ means 'implies' and $\dashv$ means 'is implied by'.

Based on these concepts, the following theorem expresses the isomorphism of $(\mathcal{L}, \dashv)$ to the lattices of antichains and parthood distributions.

Theorem 4.1. *For all $n \in \mathbb{N}$: $(\mathcal{L}_n, \dashv)$ is isomorphic to $(\mathcal{A}_n, \preceq)$ and $(\mathcal{B}_n, \sqsubseteq)$.*

Proof. See electronic supplementary material, appendix §2. ∎

[1]There is a slight ambiguity in the definition of Ψ since the *order* of the conjunctions $\bigwedge_{i \in a} \phi_i$ and statements ϕ_i is not specified. This problem can be solved, however, by choosing any enumeration of the elements $\mathbf{a}$ of the powerset of $\{1, \ldots, n\}$ and ordering the conjunctions $\bigwedge_{i \in \mathbf{a}} \phi_i$ accordingly. The propositional variables ϕ_i within the conjunctions may simply be ordered by ascending order of their indices.

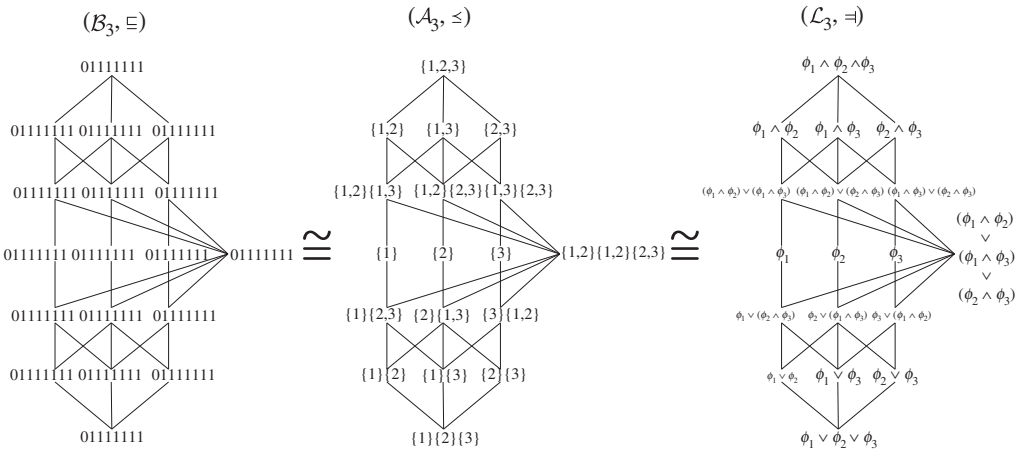

Figure 7. The three isomorphic worlds of partial information decomposition: parthood distributions, antichains and logical statements.

Corollary 4.2. *For all $n \in \mathbb{N}$: $(\mathcal{L}_n, \dashv)$ is a poset and specifically a lattice.*

In this way, the logical perspective is put on equal footing with the parthood perspective and 'antichain' perspective described by Williams & Beer [13]. They are in fact three equivalent ways to describe the mathematical structure underlying PID. These three 'worlds' of PID are illustrated in figure 7 for the case of three information sources.

Intuitively, the logic lattice can be understood as a hierarchy of logical constraints describing how (i.e. via which collections of sources) information about the target may be accessed. The information atom associated with a node $\tilde{\alpha}$ in the logic lattice is *exactly* the information about the target that can be accessed in the way described by the constraint $\tilde{\alpha}$. For example, the information shared by all sources $\Pi(\{1\}, \{2\}, \{3\})$ is to be found at the very bottom of the logic lattice because access to this information is constrained in the least possible way: the shared information can be accessed via *any* source (i.e. via source 1 *or* source 2 *or* source 3). By monotonicity, the shared information is of course also accessible via any *collection* of sources so that in total there are seven ways to access it (one per collection). By contrast, the all-way synergy $\Pi(\{1, 2, 3\})$ is located at the very top of the logic lattice because access to it is most heavily constrained: the synergy can only be accessed if all sources are known at the same time. Hence, there is only a single way (collection) to access it. All other atoms are to be found in between these two extremes. For instance, the atom corresponding to the constraint $\phi_1 \vee (\phi_2 \wedge \phi_3)$ is exactly the information that can be accessed either via source 1 or via sources 2 and 3 jointly (and of course via any superset of these collections by monotonicity) *but not in any other way* (i.e. not via sources 2 or 3 individually). So in total there are five ways to access it corresponding to the collections $\{1\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}$. In general, the atoms on the k-th level of the logic lattice (starting to count at the top) are precisely the atoms that can be accessed via k collections of sources (compare this to the very similar insight in §2c(iii)).

Finally, one may also associate a redundant information term with each node in the logic lattice by interpreting the statements as *sufficient conditions* for access instead of *constraints*, i.e. sufficient and necessary conditions, on access. For instance, the redundancy associated with the statement $\phi_1 \wedge \phi_2 \wedge \phi_3$ would be all information for which joint knowledge of all three sources is sufficient. But this is of course *all* information contained in the sources, i.e. the entire joint mutual information. By contrast, the information atom associated with the same statement is the information for which joint knowledge of all three sources is not only sufficient but also necessary, i.e. it cannot be obtained via any other collection of sources. Or put generally: while the redundancy is the information we obtain *if* we have knowledge of certain collections of sources, the information atom is the information we obtain *if and only if* we have such knowledge. Defined

in this way, the redundant information of a lattice node is again the sum of atoms associated with nodes below and including it.

In this way, the logical perspective can be used as an independent starting point to develop PID theory. Instead of characterizing atoms by their defining parthood relations one might equally characterize them by their defining access constraints and relate them to the notion of redundant information in the way just described. This is summarized in the following core principle:

Principle 4.3. Each atom of information is characterized by a logical constraint describing via which collections of sources it can be accessed. The atom $\Pi(\tilde{\alpha})$ associated with constraint $\tilde{\alpha} = \bigvee_{\mathbf{a}\in\alpha} \bigwedge_{i\in\mathbf{a}} \phi_i$ is exactly that part of the joint mutual information about the target that can be accessed if and only if we have knowledge of any one of the collections of sources **a**.

Now that we have fully introduced both the parthood and logical approaches to PID it is worth noting their key difference to the original 'antichain' approach by Williams and Beer: whereas the parthood and logic approaches are looking at the problem from the perspective of the atoms and seek to describe their defining parthood relations/access constraints, the antichain-based approach starts off by placing certain axioms on measures of redundant information leading to the insight that the definition of redundancy may be restricted to antichains. The atoms are then *indirectly* introduced in terms of a Moebius inversion over the lattice of antichains.

The next section highlights an additional use of logic lattices, namely as a mathematical tool to analyse the structure of PID lattices.

(b) Using logic lattices as a mathematical tool to analyse the structure of partial information decomposition lattices

One advantage that logic lattices have over the lattices of antichains and parthood distributions is that their ordering relationship is particularly natural and well understood: logical implication between statements. By contrast, the ordering relation $\preceq$ on the lattice of antichains only seems to have been studied in quite restricted order theoretic contexts so far. Furthermore, it is a purely technical concept that does not have a clear-cut counterpart in ordinary language. Because of the simplicity of its ordering relation, many important order theoretic concepts have a simple interpretation within the logic lattice. This makes it a useful tool to understand the structure of the lattice itself which in turn is relevant to the computation of information atoms.

There is an interesting fact about the statements in $\mathcal{L}$ that will be useful in the following investigations: they correspond to statements with monotonic truth-tables. The truth-table $T_{\tilde{\alpha}} : \mathcal{V} \to \{0, 1\}$ of a statement $\tilde{\alpha}$ describes which models $V \in \mathcal{V}$ satisfy $\tilde{\alpha}$ ('make $\tilde{\alpha}$ true'), i.e.

$$T_{\tilde{\alpha}}(V) = \begin{cases} 1 & \text{if } \models_V \tilde{\alpha} \\ 0 & \text{otherwise.} \end{cases} \tag{4.2}$$

A truth-table T shall be called *monotonic* just in case $\forall i \in \{1, \ldots, n\}$

$$(V(\phi_i) = 1 \to V'(\phi_i) = 1) \Rightarrow \big(T(V) = 1 \to T(V') = 1\big). \tag{4.3}$$

In other words, suppose a statement $\tilde{\alpha}$ is satisfied by a valuation V. Now suppose further that a new valuation V' is constructed by flipping one or more zeros to one in V. Then $\tilde{\alpha}$ has to be satisfied by V' as well. Making some ϕ_i true that were previously false cannot make $\tilde{\alpha}$ false if it was previously true. With this terminology at hand the following proposition can be formulated:

Proposition 4.4. *All $\tilde{\alpha} \in \mathcal{L}$ have monotonic truth-tables. Conversely, for all monotonic truth-tables T, there is exactly one $\tilde{\alpha} \in \mathcal{L}$ such that $T_{\tilde{\alpha}} = T$. In other words, the statements in $\mathcal{L}$ are, up to logical equivalence, exactly the statements of propositional logic with monotonic truth-tables.*

Proof. See electronic supplementary material, appendix §3.1 ■

Now, it was shown in [14] that the information atoms have a closed form solution in terms of the *meets* of any subset of children of the corresponding node in the lattice. The *meet* (infimum)

and *join* (supremum) operations, however, have quite straightforward interpretations on $(\mathcal{L}, \dashv)$: The meet of two statements $\tilde{\alpha}$ and $\tilde{\beta}$ is the strongest statement logically weaker than both $\tilde{\alpha}$ and $\tilde{\beta}$. Similarly, the join is the weakest statement logically stronger than both $\tilde{\alpha}$ and $\tilde{\beta}$. The meet is logically equivalent (though not identical) to the *disjunction* of $\tilde{\alpha}$ and $\tilde{\beta}$ while the join is logically equivalent (though not identical) to their *conjunction*. The conjunction and disjunction of two elements of $\mathcal{L}$ do generally not lie in $\mathcal{L}$ because they do not necessarily have the appropriate form (disjunction of logically independent conjunctions). However, this can easily be remedied because both the disjunction and the conjunction of elements of $\mathcal{L}$ have monotonic truth-tables. Thus, by proposition 4.4, there is a unique element in $\mathcal{L}$ with the same truth-table in both cases. These elements are therefore the meet and join. The detailed construction of meet and join operators is presented in electronic supplementary material, appendix §3.5.

Let us now turn to the notions of child and parent. A *child* of a statement $\tilde{\alpha} \in \mathcal{L}$ is a strongest statement strictly weaker than $\tilde{\alpha}$. Similarly, a *parent* of $\tilde{\alpha}$ is a weakest statement strictly stronger than $\tilde{\alpha}$. The following three propositions provide, first, a characterization of children in terms of their truth tables, second, a lower bound on the number of children of a statement, and third, an algorithm to determine all children of a statement. Owing to the isomorphism of antichains, parthood distributions, and logical statements, the propositions can be used to study any of these three structures.

Proposition 4.5 (Characterization of children). *$\tilde{\gamma} \in \mathcal{L}$ is a direct child of $\tilde{\alpha} \in \mathcal{L}$ if and only if $\tilde{\gamma}$ is true in all cases in which $\tilde{\alpha}$ is true plus exactly one additional case, i.e. just in case $T_{\tilde{\alpha}}(V) = 1 \rightarrow T_{\tilde{\gamma}}(V) = 1$ and $!\exists V \in \mathcal{V} : T_{\tilde{\gamma}}(V) = 1 \wedge T_{\tilde{\alpha}}(V) = 0$.*

Proof. See electronic supplementary material, appendix §3.2. ■

Proposition 4.6 (Lower bound on number of children). *Any $\alpha \in \mathcal{A}$ such that there is at least one $\mathbf{a} \in \alpha$ with $|\mathbf{a}| = k \geq 1$ has at least k children.*

Proof. See electronic supplementary material, appendix §3.3. ■

Proposition 4.7 (Algorithm to determine children). *The children of a statement $\tilde{\alpha}$ can be determined via the following algorithm (for a pseudocode version see electronic supplementary material, appendix §3.4). It proceeds in three steps:*

(i) Set k to the maximal number of ones occurring in a valuation that does not satisfy $\tilde{\alpha}$.
(i) For each valuation V that does not satisfy $\tilde{\alpha}$ and contains k ones do the following:

(a) Check if there is a valuation with k+1 ones that does not satisfy $\tilde{\alpha}$ and results from flipping one or multiple zeros in V to one, i.e. a model V' such that $V(\phi_i) = 1 \rightarrow V'(\phi_i) = 1$. If there is such a valuation, then skip step b). Otherwise, proceed.
(b) Create a new monotonic truth-table by setting V to one, otherwise leaving the truth-table of $\tilde{\alpha}$ unchanged. The statement corresponding to this truth-table is a child of $\tilde{\alpha}$.

(iii) If $k > 0$, decrease k by 1 and repeat Step 2. Otherwise, terminate.

Proof. See electronic supplementary material, appendix §3.4 ■

This concludes our discussion of the relationship between formal logic and PID. In the next section, we return to the parthood side of our story. In particular, we will address an apparent arbitrariness in the argument presented in §2c. Here we showed that the sizes of the atoms of information can be obtained once a measure of redundant information is specified. Now, one may ask of course: why redundant information? Could not the same purpose be achieved by using some other informational quantity such as synergistic or unique information? We will now discuss how the parthood approach can help answering this question in a systematic way.

5. Non-redundancy-based partial information decompositions

Let us briefly revisit the structure of the argument in §2c. It involved three steps (presented in slightly different order above): first, based on the very concept of redundant information, we phrased a condition describing which atoms are part of which redundancies (principle 2.4). Secondly, we showed that this parthood criterion entails a number of constraints on the measure $I_\cap$. Finally, we showed that, as long as these constraints are satisfied, we obtain a unique solution for the atoms of information. There is actually a fourth step: we would have to check that the information decomposition satisfies the consistency equations relating atoms to mutual information terms (equation (2.5)). However, in the case of redundant information, this condition is trivially satisfied due to the self-redundancy property. In other words, the consistency equations are themselves part of the system of equations used to solve for the information atoms.

In order to obtain an information decomposition based on a quantity other than redundant information, let's call it $I^*(T : \mathbf{a}_1, \ldots, \mathbf{a}_m)$, we may use precisely the same scheme.

(i) Define a condition $\mathcal{C}(f : \mathbf{a}_1, \ldots, \mathbf{a}_m)$ on parthood distributions f describing which atoms $\Pi(f)$ are part of $I^*(T : \mathbf{a}_1, \ldots, \mathbf{a}_m)$ for any given tuple of collections of sources $\mathbf{a}_1, \ldots, \mathbf{a}_m$. This leads to a system of equations

$$I^*(T : \mathbf{a}_1, \ldots, \mathbf{a}_m) = \sum_{\mathcal{C}(f:\mathbf{a}_1,\ldots,\mathbf{a}_m)} \Pi(f). \tag{5.1}$$

(ii) Analyse which constraints on $I^*(T : \mathbf{a}_1, \ldots, \mathbf{a}_m)$ (e.g. symmetry, idempotency, ...) are implied by this relationship.

(iii) Show that given a choice of $I^*(T : \mathbf{a}_1, \ldots, \mathbf{a}_m)$ that satisfies the constraints, a unique solution for all information atoms $\Pi(f)$ can be obtained.

(iv) Show that the solution satisfies the consistency equation (2.5) relating information atoms and mutual information terms.

Let us work through these steps in specific cases.

(a) Restricted information partial information decomposition

Recall that the redundant information of multiple collections of sources is the information we obtain *if* we have access to any of the collections. Similarly, we can define the information 'restricted by' collections of sources $\mathbf{a}_1, \ldots, \mathbf{a}_m$ as any information we obtain *only if* we have access to at least one of the collections. For instance, assuming $n = 2$, the information restricted by the first source consists of its unique information and its synergy with the second source. Both of these quantities can only be obtained if we have access to the first source.

Thus, in general the restricted information $I_{\text{res}}(T : \mathbf{a}_1, \ldots, \mathbf{a}_m)$ should consist of all the atoms that are *only* part of the information carried by some of the $\mathbf{a}_i$ *but not part of the information provided by any other collection of sources*. Thus the parthood condition $\mathcal{C}_{\text{res}}$ is given by

$$\mathcal{C}_{\text{res}}(f : \mathbf{a}_1, \ldots, \mathbf{a}_m) \Leftrightarrow (f(\mathbf{b}) = 1 \rightarrow \exists i : \mathbf{b} \supseteq \mathbf{a}_i), \tag{5.2}$$

and we obtain the relation

$$I_{\text{res}}(T : \mathbf{a}_1, \ldots, \mathbf{a}_m) = \sum_{\mathcal{C}_{\text{res}}(f:\mathbf{a}_1,\ldots,\mathbf{a}_m)} \Pi(f). \tag{5.3}$$

Just as in the case of redundant information, this relationship implies a number of invariance properties for I_{res}: it has to be symmetric, idempotent, and invariant under superset removal/addition allowing us again to restrict ourselves to the set of antichains. The analogue of the 'self-redundancy' property is that the restricted information of a collection of singletons $I_{\text{res}}(T : \{i_1\}, \ldots, \{i_m\})$ is equal to the *conditional mutual information provided by their union*

$\alpha_\cup = \bigcup_{j=1}^{m}\{i_j\}$ *conditioned on all other sources*. So if $\alpha = \{\{i_1\}, \ldots, \{i_m\}\}$ is a collection of singletons, then:

$$I_{\text{res}}(T : \alpha) = I\left(T : (S_i)_{i\in\alpha_\cup} | (S_j)_{j\in\alpha_\cup^C}\right). \tag{5.4}$$

This can be established using the chain rule for mutual information as detailed in electronic supplementary material, appendix §4.1. The next step is to show that we may obtain a unique solution for the information atoms once a measure of restricted information satisfying these conditions is given. This can be achieved in much the same way as for redundant information. The restricted information associated with an antichain α can be expressed as a sum of information atoms $\Pi(\beta)$ below and including α in a specific lattice of antichains $(\mathcal{A}, \preceq')$. This lattice is simply the dual (inverted version) of the antichain lattice $(\mathcal{A}, \preceq)$, i.e.

$$\alpha \preceq' \beta \Leftrightarrow \beta \preceq \alpha. \tag{5.5}$$

Accordingly, a unique solution is guaranteed via Moebius inversion of the relationship

$$I_{\text{res}}(T : \alpha) = \sum_{\beta \preceq' \alpha} \Pi_{\text{res}}(\alpha). \tag{5.6}$$

As a final step, we need to show that the resulting atoms stand in the appropriate relationships to mutual information terms. These relationships are given by the consistency equation (2.5). Again using the chain rule it can be shown that this equation is equivalent to a condition relating conditional mutual information to the information atoms

$$I(T : \mathbf{a}) = \sum_{f(\mathbf{a})=1} \Pi(f) \Leftrightarrow I(T : \mathbf{a}|\mathbf{a}^C) = \sum_{f(\mathbf{a}^C)=0} \Pi(f). \tag{5.7}$$

Now consider any collection of source indices $\mathbf{a} = \{j_1, \ldots, j_m\}$, then we obtain

$$I\left(T : \mathbf{a}|\mathbf{a}^C\right) \overset{\text{Eq.(5.4)}}{=} I_{\text{res}}(T : \{j_1\}, \ldots, \{j_m\}) \tag{5.8}$$

$$\overset{\text{Eq.(5.3)}}{=} \sum_{f(\mathbf{b})=1 \to \exists i : \mathbf{b} \supseteq \{j_i\}} \Pi_{\text{res}}(f) \tag{5.9}$$

$$= \sum_{f(\mathbf{a}^C)=0} \Pi_{\text{res}}(f), \tag{5.10}$$

where the last equality follows because in the case of singletons the parthood condition $\mathcal{C}_{\text{res}}$ reduces to $f(\alpha_\cup^C) = 0$. This establishes that the resulting atoms satisfy the consistency condition and we obtain a valid PID. In the following section, we will use the same approach to analyse the question of whether a synergy-based PID is possible.

(b) Synergy-based partial information decomposition

Note that the restricted information of multiple collections of sources stands in a direct correspondence to a weak form of synergy which we will denote by $I_{\text{ws}}(T : \mathbf{a}_1, \ldots, \mathbf{a}_m)$. This quantity is to be understood as *the information about the target we cannot obtain from any individual collection* $\mathbf{a}_i$. Accordingly, the parthood criterion is

$$\mathcal{C}_{\text{ws}}(f : \mathbf{a}_1, \ldots, \mathbf{a}_m) \Leftrightarrow (\forall i \in \{1, \ldots, m\} : f(\mathbf{a}_i) = 0). \tag{5.11}$$

But this information is of course the same as the information that we can get only if some *other* collections are known (except subcollections of course), i.e.

$$I_{\text{ws}}(T : \mathbf{a}_1, \ldots, \mathbf{a}_m) = I_{\text{res}}(T : (\mathbf{b} \mid \forall i \mathbf{b} \not\subseteq \mathbf{a}_i)). \tag{5.12}$$

Consider the case of two sources: the information we cannot get from source 2 alone, $I_{\text{ws}}(T : \{2\})$, is the same as the information we can get only if the first source is known, $I_{\text{res}}(T : \{1\})$: unique information of source 1 plus synergistic information.

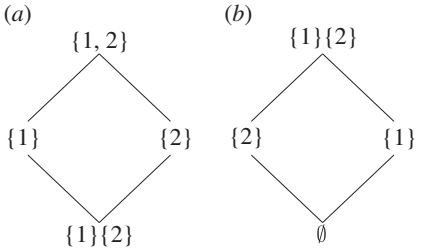

Figure 8. (*a*) Antichain lattice $(\mathcal{A}_2, \preceq)$ for two sources. Summing up the atoms *above* and including a node yields the restricted information of that node. (*b*) Extended constraint lattice for two sources. The weak synergy associated with a node in the extended constraint lattice is the sum of atoms above and including the corresponding node in the left lattice. Note that following a widespread convention we left out the outer curly brackets around the antichains.

Owing to this correspondence, the argument presented above can also be used to show that a consistent PID can be obtained by fixing a measure I_{ws} of weak synergy. Once such a measure is given we can first translate it to the corresponding restricted information terms and then perform the Moebius inversion of equation (5.6) (alternatively, the above argument could be redeveloped directly for I_{ws} with minor modifications)

Interestingly, if we associate with every antichain α in the lattice $(\mathcal{A}, \preceq)$ the corresponding $I_{\mathrm{ws}}(T:\beta)$ (so that $I_{\mathrm{res}}(T:\alpha) = I_{\mathrm{ws}}(T:\beta)$), then the β form an isomorphic lattice but with a different ordering (figure 8). Just as the original antichain lattice this structure on the antichains has been introduced by Crampton & Loizou [19].

In the PID field, a restricted version of this lattice (i.e. restricted to a certain subset of antichains) has been described by [24,25] under the name 'constraint lattice'. This terms is also appropriate in the present context: intuitively, if we move up the constraint lattice we encounter information that satisfies more and more constraints. First, all of the information in the sources ($I_{\mathrm{ws}}(T:\emptyset)$). This is the case of no constraints. Then all the information that is not contained in a particular individual source ($I_{\mathrm{ws}}(T:\{1\})$ and $I_{\mathrm{ws}}(T:\{2\})$). And finally, the information that is not contained in any individual source ($I_{\mathrm{ws}}(T:\{1\},\{2\})$) .

Most recently, the full version of the lattice (i.e. defined on all antichains) has been used by [26] to formulate a synergy centred information decomposition. They call the lattice *extended constraint lattice* and define 'synergy atoms' S_∂ in terms of a Moebius inversion over it. The concept of synergy $S^{\boldsymbol{\alpha}}$ used in this approach closely resembles what we have called weak synergy. However, the decomposition is *structurally different* from the type of decomposition discussed here and generally assumed in previous work on PID. Even though it leads to the same number of atoms, these atoms do not stand in the expected relationships to mutual information. For instance, in the 2-sources case, there is no pair of atoms that necessarily adds up to the mutual information provided by the first source and no such pair of atoms for the second source. The consistency equation (2.5) is not satisfied (except for the full set of sources). This means that synergy atoms S_∂ are not directly comparable to standard PID atoms Π. They represent different types of information.

Let us now move towards stronger concepts of synergistic information. The reason for the term 'weak' synergy is that a key ingredient of synergy seems to be missing in its definition: intuitively, the synergy of multiple sources is the information that cannot be obtained from any individual source but that becomes 'visible' once we know all the sources at the same time. However, the definition of weak synergy only comprises the first part of this idea. The weak synergy $I_{\mathrm{ws}}(T:\mathbf{a}_1,\ldots,\mathbf{a}_m)$ also contains parts that do not become visible even if we have access to *all* $\mathbf{a}_i$. For instance, given $n=3$, the weak synergy $I_{\mathrm{ws}}(T:\{1\},\{2\})$ also contains the unique information of the third source $\Pi(\{3\})$ because this quantity is accessible from neither the first nor the second source.

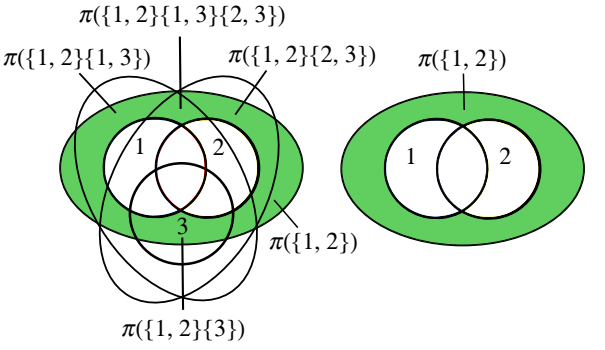

Figure 9. Geometrical interpretation of moderate synergy $I_{\text{ms}}(T : \{1\}, \{2\})$ for 2 and 3 sources.

So let us add this missing ingredient by strengthening the parthood criterion

$$\mathcal{C}_{\text{ms}}(f : \mathbf{a}_1, \ldots, \mathbf{a}_m) \Leftrightarrow (\forall i \in \{1, \ldots, m\} : f(\mathbf{a}_i) = 0 \;\&\; f(\alpha_\cup) = 1). \tag{5.13}$$

We obtain a moderate type of synergy we denote by $I_{\text{ms}}(T : \mathbf{a}_1, \ldots \mathbf{a}_m)$. It has a nice geometrical interpretation: in an information diagram it corresponds to all atoms *outside* of all areas associated with the mutual information carried by some $\mathbf{a}_i$ but *inside* the area associated with the mutual information carried by the union of the $\mathbf{a}_i$ (figure 9). Furthermore, we can immediately see that the parthood condition cannot be satisfied for individual collections $\mathbf{a}$ (it demands $f(\mathbf{a}) = 0$ and $f(\mathbf{a}) = 1$ at the same time). This makes intuitive sense because the synergy of an individual collection appears to be an ill-defined concept: at least two things have to come together for there to be synergy. We will get back to the case of individual collections below.

Let us first see what properties are implied by $\mathcal{C}_{\text{ms}}$. It can readily be shown that I_{ms} is symmetric, idempotent and invariant under *subset* removal. This again allows us to restrict the domain of I_{ms} to the antichains. Additionally, I_{ms} satisfies the following condition:

$$\text{If } \exists i : \alpha_\cup = \mathbf{a}_i, \text{ then } I_{\text{ms}}(T : \alpha) = 0 \textbf{ (zero condition)}. \tag{5.14}$$

This property says that whenever the union of the collection happens to be equal to one of collections then the moderate synergy must be zero. This is in particular the case for the moderate 'self-synergy' of a single collection. On first sight, this raises a problem since the synergy equations associated with individual collections become trivial ($0 = 0$) and do not impose any constraints on the atoms. This situation can be remedied, however, by noting that these missing constraints are provided by the consistency equations relating the atoms to mutual information/conditional mutual information. In this way, a unique solution for the atoms is indeed guaranteed (one could also axiomatically set the 'self-synergies' to the respective conditional mutual information terms). The proof of this statement is given in electronic supplementary material, appendix §4.2.

An instructive fact about the moderate synergy-based PID is that the underlying system of equations does not have the structure of a Moebius inversion over a lattice: there is no arrangement of atoms into a lattice such that each $I_{\text{ms}}(T : \alpha)$ turns out to be the sum of atoms below and including a particular lattice node. The reason is that any finite lattice always has a unique least element. In other words, some atom would have to appear at the very bottom of the lattice and would therefore be contained in all synergy terms. However, in the case of moderate synergy, there is no such atom for $n \geq 3$. The only viable candidate would be the overall synergy $\Pi(\{1, \ldots, n\})$. But due to the condition that the synergistic information has to become visible if we know all collections in question, this atom is not contained e.g. in $I_{\text{ms}}(T : \{1\}, \{2\})$.

Now one may wonder if the concept of synergy can be strengthened even further by demanding that the synergistic information should not be accessible from the *union of any proper subset* of the collections in question. For instance, the synergistic information $I_{\text{syn}}(T : \{1\}\{2\}\{3\})$ of sources 1, 2 and 3 should not be accessible from the collections $\{1, 2\}$, $\{1, 3\}$, or $\{2, 3\}$. We have to

know *all three sources* to get access to their synergy. Thus, we may add this third constraint to obtain a strong notion of synergy we denote by $I_{\text{syn}}(T : \mathbf{a}_1, \ldots, \mathbf{a}_m)$. An atom $\Pi(f)$ should satisfy the corresponding parthood condition $\mathcal{C}_{\text{syn}}(f : \mathbf{a}_1, \ldots, \mathbf{a}_m)$ just in case

(i) $f(\bigcup_{i=1}^{m} \mathbf{a}_i) = 1$
(ii) $\forall i \in \{1, \ldots, m\} : f(\mathbf{a}_i) = 0$
(iii) $\forall J \subset \{1, \ldots, m\}, |J| \geq 2 : \bigcup_{j \in J} \mathbf{a}_j \neq \bigcup_{i=1}^{m} \mathbf{a}_i \rightarrow f(\bigcup_{j \in J} \mathbf{a}_j) = 0.$

The last condition is phrased as a conditional because the union of a proper subset of collection might happen to be equal to the union of all collections in question. Consider the case of three sources and the synergy $I_{\text{syn}}(T : \{1,2\}\{1,3\}\{2,3\})$. In this case, the union of a proper subset of these collections, for instance $\{1,2\} \cup \{1,3\}$, happens to be equal to the union of all $\mathbf{a}_i$.

Unfortunately, we do not obtain enough linearly independent equations to uniquely determine the atoms of information. This can be shown using the example of three sources. According to the parthood criterion, $I_{\text{syn}}(T : \{1\}\{2\}\{3\}) = \Pi(\{1,2,3\})$. But also $I_{\text{syn}}(T : \{1,2\}\{1,3\}\{2,3\}) = \Pi(\{1,2,3\})$. This means that we do not obtain independent equations for each antichain. Or in linear algebras terms: our coefficient matrix will have two linearly dependent (actually identical) rows. Thus, a measure of strong synergy as described by $\mathcal{C}_{\text{syn}}$ cannot induce a unique PID.

(c) Unique information partial information decomposition

Let us briefly discuss the last obvious candidate quantity for determining the PID atoms: unique information [27]. The appropriate parthood criterion for a measure of unique information I_{unq} seems straightforward in the case of individual collections $\mathbf{a}$: it should consist of all atoms that are part of the information provided by the collection $\mathbf{a}$ but not part of the information provided by any other collection. This is what makes this information 'unique' to the collection. Since there is always just one such atom this means that $I_{\text{unq}}(T : \mathbf{a}) = \Pi(\mathbf{a})$. For instance, $I_{\text{unq}}(T : \{1\}) = \Pi(\{1\})$, as expected. However, defining I_{unq} only for individual collections does not yield enough equations to solve for the atoms. We need one equation per antichain/parthood distribution, and hence, some notion of the unique information associated with *multiple* collections $\mathbf{a}_1, \ldots, \mathbf{a}_m$. This is a trickier question. What does it mean for information to be unique to these collections? Certainly, uniqueness demands that this information should not be contained in any *other* collection. But what about the collections $\mathbf{a}_1, \ldots, \mathbf{a}_m$ themselves? It seems that the appropriate condition is that the unique information should consist of atoms that are contained in *all* of these collections. This idea aligns well with ordinary language: for instance, saying that a certain protein is unique to sheep and goats means that this protein is found in *both sheep and goats and nowhere else*. Using this idea, the parthood criterion becomes

$$\mathcal{C}_{\text{unq}}(f : \mathbf{a}_1, \ldots, \mathbf{a}_m) \Leftrightarrow \left(f(\mathbf{a}) = 1 \leftrightarrow \exists i : \mathbf{a} \supseteq \mathbf{a}_i\right). \tag{5.15}$$

However, this condition simply defines the atom $\Pi(\mathbf{a}_1, \ldots, \mathbf{a}_m)$ making the unique information based PID possible but maybe not very helpful: it just amounts to defining all the atoms separately because $I_{\text{unq}}(T : \alpha) = \Pi(\alpha)$ for all antichains α.

6. Parthood descriptions versus quantitative descriptions

Before concluding we would like to briefly point out an issue that arises quite naturally when thinking about information theory from a parthood perspective and that merits a few remarks: throughout this paper, we have drawn a distinction between *parthood* relationships and *quantitative* relationships between information contributions. In particular, principles 2.2 and 2.4 express parthood relationships between information atoms on the one hand and mutual information/redundant information on the other. Principle 2.3 by contrast describes the quantitative relationship between any information contribution and the parts it consists of. It is crucial to draw this distinction because these principles are logically independent. Consider the

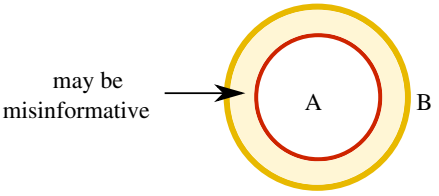

Figure 10. Illustration of the idea that the information provided by a logically weaker statement A is always *part of* the information of a stronger statement B, even though the latter may provide *less bits* of information. This phenomenon can be explained in terms of the misinformative, i.e. negative, contribution of the surplus information provided by B (the shaded ring). (Online version in colour.)

case of two sources: In this case, one could agree that the joint mutual information should consist of four parts while disagreeing that it should be the *sum* of these parts. On other hand, one could agree that the joint mutual information should be the sum of its parts but disagree that it consists of four parts.

The distinction between parthood relations and quantitative relations is also important in the argument that the redundant information provided by multiple statements is the information carried by the truth of their disjunction. One of the two motivations for this idea was based on the principle that the information provided by a statement A is always *part of* the information provided by any stronger statement B. This does not mean however, that statement A necessarily provides *quantitatively* less information than B (i.e. *less bits* of information). In fact, this latter principle would contradict classical information theory. Here is why: suppose the pointwise mutual information $i(t:s) = i(t:S=s)$ is negative. Now, consider any *tautology* such as $S=s \vee \neg(S=s)$. Certainly, this statement is *logically weaker* than $S=s$ because a tautology is implied by any other statements. Furthermore, the probability of the tautology being true is equal to 1. Therefore, the information $i(t:S=s \vee \neg(S=s))$ provided by it is equal to 0. But this means $i(t:S=s) < i(t:S=s \vee \neg(S=s))$ even though $S=s \vee \neg(S=s) \dashv S=s$.

Nonetheless, there certainly is a sense in which a stronger statement B provides 'more' information than a weaker statement A: the information provided by A is *part of* the information provided by B. If we know B is true than we can by assumption infer that A is true, and hence, we have access to all the information provided by A. The fact that the stronger statement B may nonetheless provide less bits of information can be explained in terms of misinformation: if we know B is true, then we obtain all the information carried by A *plus some additional information*. If it happens that this surplus information is misinformative, i.e. negative, then quantitatively B will provide less information than A. This idea is illustrated in figure 10.

Importantly, the possible negativity and non-monotonicity of $i_\cap^{\mathrm{sx}}$ as well as the potential negativity of π^{sx} can be *completely* explained in terms of misinformative contributions in the following sense: it is possible [28] to uniquely separate $i_\cap^{\mathrm{sx}}$ into an informative part $i_\cap^{\mathrm{sx+}}$ and a misinformative part $i_\cap^{\mathrm{sx-}}$ such that

$$i_\cap^{\mathrm{sx}}(t:\alpha) = i_\cap^{\mathrm{sx+}}(t:\alpha) - i_\cap^{\mathrm{sx-}}(t:\alpha). \tag{6.1}$$

Now, each of these components can be shown to be non-negative and monotonically increasing over the lattice. Moreover, the induced informative and misinformative atoms $\pi^{\mathrm{sx+}}$ and $\pi^{\mathrm{sx-}}$ are non-negative as well [20]. In other words, once we sperate out informative and misinformative components any violations of non-negativity and monotonicity disappear. Hence, these violations can be fully accounted for in terms of misinformative contributions.

7. Conclusion

In this paper, we connected PID theory with ideas from mereology, i.e. the study of parthood relations, and formal logic. The main insights derived from these ideas are that the general

structure of information decomposition as originally introduced by Williams & Beer [13] can be derived entirely from (i) parthood relations between information contributions and (ii) in terms of a hierarchy of logical constraints on how information about the target can be accessed. In this way, the theory is set up from the perspective of the atoms of information, i.e. the quantities we are ultimately interested in. The *n*-sources PID problem has conventionally been approached by defining a measure of redundant information which in turn implies a unique solution for the atoms of information. We showed how such a measure can be defined in terms of the information provided by logical statements of a specific form. We discussed furthermore how the parthood perspective can be used to systematically address the question of whether a PID may be determined based on concepts other than redundancy. In doing so, we showed that this is indeed possible in terms of measures of 'restricted information', 'weak synergy', and 'moderate synergy' but not in terms of 'strong synergy'. We hope to have shown that there are deep connections between mereology, formal logic and information decomposition that future research in these fields may benefit from.

Data accessibility. This article has no additional data.

Authors' contributions. A.G. conceived the parthood-based and logic-based formulations of PID and wrote the original manuscript except for the introduction which was provided by M.W. M.W. originally conceived the $i_\cap^{sx}$ measure of redundant information rederived in §3. A.M. provided critical feedback regarding the mathematical aspects of the paper. All authors were involved in revising the manuscript and refining the ideas presented therein. All authors gave final approval for publication and agree to be held accountable for the work performed therein.

Competing interests. We declare we have no competing interests.

Funding. M.W., A.M. and A.G. are employed at the Campus Institute for Dynamics of Biological Networks (CIDBN) funded by the Volkswagen Stiftung. M.W. and A.M. received support from the Volkswagenstiftung under the programme 'Big Data in den Lebenswissenschaften'. This work was supported by a funding from the Ministry for Science and Education of Lower Saxony and the Volkswagen Foundation through the 'Niedersächsisches Vorab'.

Acknowledgements. We thank Kyle Schick-Poland, David Ehrlich and Andreas Schneider for helpful comments on the draft.

References

1. Schneidman E, Bialek W, Berry MJ. 2003 Synergy, redundancy, and independence in population codes. *J. Neurosci.* **23**, 11 539–11 553. (doi:10.1523/JNEUROSCI.23-37-11539.2003)
2. McGill W. 1954 Multivariate information transmission. *Trans. IRE Professional Group Information Theory* **4**, 93–111.
3. MacKay DJC, Mac Kay DJC. 2003 *Information theory, inference and learning algorithms*. Cambridge, UK: Cambridge university press.
4. Wibral M, Finn C, Wollstadt P, Lizier JT, Priesemann V. 2015 Bits from brains for biologically inspired computing. *Front. Rob. AI* **2**, 5.
5. Rao RPN, Ballard DH. 1999 Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* **2**, 79–87. (doi:10.1038/4580)
6. Kay JW, Phillips WA. 2011 Coherent Infomax as a computational goal for neural systems. *Bull. Math. Biol.* **73**, 344–372. (doi:10.1007/s11538-010-9564-x)
7. Wibral M, Priesemann V, Kay JW, Lizier JT, Phillips WA. 2017 Partial information decomposition as a unified approach to the specification of neural goal functions. *Brain Cogn.* **112**, 25–38. (doi:10.1016/j.bandc.2015.09.004)
8. Rauh J. 2017 Secret sharing and shared information. *Entropy* **19**, 601. (doi:10.3390/e19110601)
9. Lizier JT, Flecker B, Williams PL. 2013 Towards a synergy-based approach to measuring information modification. In *2013 IEEE Symp. on Artificial Life (ALIFE)*, pp. 43–51. Piscataway, NJ: IEEE.
10. Wibral M, Finn C, Wollstadt P, Lizier JT, Priesemann V. 2017 Quantifying information modification in developing neural networks via partial information decomposition. *Entropy* **19**, 494. (doi:10.3390/e19090494)

11. Rosas F, Mediano PAM, Ugarte M, Jensen HJ. 2018 An information-theoretic approach to self-organisation: emergence of complex interdependencies in coupled dynamical systems. *Entropy* **20**, 793. (doi:10.3390/e20100793)
12. Rosas FE, Mediano PAM, Jensen HJ, Seth AK, Barrett AB, Carhart-Harris RL, Bor D. 2020 Reconciling emergences: an information-theoretic approach to identify causal emergence in multivariate data. (http://arxiv.org/abs/2004.08220).
13. Williams PL, Beer RD. 2010 Nonnegative decomposition of multivariate information. (http://arxiv.org/abs/1004.2515).
14. Finn C, Lizier J. 2018 Pointwise partial information decomposition using the specificity and ambiguity lattices. *Entropy* **20**, 297. (doi:10.3390/e20040297)
15. Cover TM. 1999 *Elements of information theory*. Hoboken, NJ: John Wiley & Sons.
16. Shannon CE. 1948 A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423. (doi:10.1002/j.1538-7305.1948.tb01338.x)
17. Stanley RP. 1997 Enumerative combinatorics, vol. 1. 1997. Cambridge Studies in Advanced Mathematics. Cambridge, UK: Cambridge University Press.
18. Tittmann P. 2014 *Einführung in die Kombinatorik*. New York, NY: Springer.
19. Crampton J, Loizou G. 2000 *Two partial orders on the set of antichains*. Research note, September.
20. Makkeh A, Gutknecht AJ, Wibral M. 2020 Introducing a differentiable measure of pointwise shared information. *Phys. Rev. E* **103**, 032149. (doi:10.1103/PhysRevE.103.032149)
21. Fano RM. 1961 *Transmission of information: a statistical theory of communication mit press*. New York, NY: Cambridge.
22. Wollstadt P, Lizier JT, Vicente R, Finn C, Martínez-Zarzuela M, Mediano P, Novelli L, Wibral M. 2017 IDTxl: the information dynamics Toolkit xl: a Python package for the efficient analysis of multivariate information dynamics in networks (http://arxiv.org/abs/1807.10459).
23. Smullyan RM. 1995 *First-order logic*. Mineola, NY: Dover Publications.
24. Ay N, Polani D, Virgo N. 2019 Information decomposition based on cooperative game theory. (http://arxiv.org/abs/1910.05979).
25. James RG, Emenheiser J, Crutchfield JP. 2018 Unique information via dependency constraints. *J. Phys. A: Math. Theor.* **52**, 014002. (doi:10.1088/1751-8121/aaed53)
26. Rosas F, Mediano P, Rassouli B, Barrett A. 2020 An operational information decomposition via synergistic disclosure. *J. Phys. A: Math. Theor.* **53**, 485001. (doi:10.1088/1751-8121/abb723)
27. Bertschinger N, Rauh J, Olbrich E, Jost J, Ay N. 2014 Quantifying unique information. *Entropy* **16**, 2161–2183. (doi:10.3390/e16042161)
28. Finn C, Lizier J. 2018 Probability mass exclusions and the directed components of mutual information. *Entropy* **20**, 826. (doi:10.3390/e20110826)