European Journal for Philosophy of Science (2024) 14:9
https://doi.org/10.1007/s13194-023-00566-1

PAPER IN THE PHILOSOPHY OF THE LIFE SCIENCES

Adding causality to the information-theoretic perspective on individuality

Pierrick Bourrat[1,2]

Received: 2 July 2023 / Accepted: 19 December 2023 / Published online: 16 February 2024
© The Author(s) 2024, corrected publication 2024

Abstract
I extend work from Krakauer et al. (2020), who propose a conception of individuality as the capacity to propagate information through time. From this conception, they develop information-theoretic measures. I identify several shortcomings with these measures—in particular, that they are associative rather than causal. I rectify this shortcoming by deriving a causal information-theoretic measure of individuality. I then illustrate how this measure can be implemented and extended in the context of evolutionary transitions in individuality.

1 Introduction

In a recent publication, Krakauer et al. (2020) proposed a novel way to define individuality in information-theoretic terms, which departs from more common ways to define this concept using an evolutionary, metabolic, or immunological perspective. They propose that an individual is an aggregate entity that propagates information through time while preserving its integrity. Starting from this general statement, they propose information-theoretic measures based on mutual information to characterize it. One intuitive way to understand their approach is to suppose a population of biological entities whose activities can be measured. The measurements obtained can then be packaged into summary measurements—to which I will refer as a particular 'coarse graining'—corresponding to candidate individuals. With this particular coarse-graining at hand, one can then assess the extent to which it is effective in tracking the state of the population at a later point in time. Levels of individuality, from this approach, correspond to coarse-grainings that permit the most accurate predictions. Krakauer et al.'s approach provides a principled way to define individuality. Further,

✉ Pierrick Bourrat
p.bourrat@gmail.com

1 Macquarie University, Department of Philosophy, North Ryde, NSW 2109, Australia

2 The University of Sydney, Department of Philosophy & Charles Perkins Centre, Camperdown, NSW 2006, Australia

the range of applications for these measures goes well beyond the notion of biological individuality. In this paper, I will primarily be concerned with their approach in a biological context and, more particularly, an evolutionary one.

In addition to providing a principled way to define individuality, Krakauer et al.'s approach has several other strengths. The first is that it anchors a conception of individuality as a property that comes in degree rather than a property that is either present or absent. This makes sense from the point of view of the history of life where new levels of individuality emerged from a succession of so-called evolutionary transitions in individuality (ETIs), a topic I will revisit in Section 6. Viewed from the perspective of ETIs, collective-level individuality is a property that is often gained gradually rather than abruptly. Second, Krakauer et al.'s approach leads to a conception of individuality devoid of metaphysical commitments. This is so because, following their account, no level is regarded as *the* level of individuality; a level is fundamentally a level of description, and there can be multiple levels at which individuality can be attributed. This permits sidestepping the problem of defining in what sense a new level of individuality can emerge. Here, 'emergence' amounts to simply recognizing that describing a setting at a particular coarse-graining permits accurate predictions of the future states of this setting (see Bourrat, 2023 for more on this point).

Notwithstanding my enthusiasm for Krakauer et al.'s approach, their account has several flaws and requires further elaboration. In this paper, I focus on one shortcoming: that the account is based on associative rather than causal measures. Using the interventionist account of causation (Pearl, 2009; Woodward, 2003) and recent work applying this account in the context of information theory (Griffiths et al., 2015; Pocheville et al., 2017; Bourrat, 2019), I refine Krakauer et al.'s approach by deriving a *causal* measure of individuality. Second, I show how this causal measure can be deployed to assess whether an ETI has occurred, following the ecological scaffolding model proposed by Black et al. (2020). Further, I provide additional causal information-theoretic measures that could be useful in the context of this model.

2 Basics of information theory

Krakauer et al.'s account is grounded in information theory (Shannon, 1948; Cover & Thomas, 2006). Thus, to assess their proposal, it is important to understand the basics of this theoretical framework. At the heart of information theory is Shannon's (1948) notion of entropy, symbolized by H. The Shannon entropy of a random variable X, $H(X)$, represents the expected uncertainty about X. Uncertainty can be conceived of here as the expected number of yes/no questions that one must ask to know the value of the variable. For instance, in the case of a variable with four equiprobable possible values, one must ask, on average, two yes/no questions to know the value of the variable. It follows that the Shannon entropy of this variable is 2 bits of information. Formally, assuming that X is a discrete random variable (an assumption I will make for all the variables used here), we can define its entropy as:

$$H(X) = -\sum_{i=1}^{N} p(x_i) \log p(x_i), \tag{1}$$

where $\sum_{j=1}^{N}$ represents the sum over the N possible values of X, $p(x_i)$ represents the probability of the value x_i occurring, and log is the logarithm operator. Any base for the logarithm can be used. Two commonly used bases are 2, which results in entropy measures in *bits*, and e, which results in entropy measures in *nats*.

Entropy can also be defined for two or more variables. This is known as 'joint entropy,' which represents the amount of uncertainty about this set of variables. Formally, the joint entropy of X and Y, $H(X, Y)$, is:

$$H(X, Y) = -\sum_{i=1}^{N}\sum_{j=1}^{M} p(x_i, y_j) \log p(x_i, y_j), \tag{2}$$

where $\sum_{j=1}^{M}$ represents the sum over the M possible values of Y, and $p(x, y)$ represents the joint probability of the values x_i and y_j occurring together.

Similarly, entropy can be defined for a variable conditioned on another. The resulting entropy, known as 'conditional entropy,' represents the amount of uncertainty about a variable that remains once the value of another variable is known. Formally, starting from the definitions of entropy and joint entropy, the entropy of Y conditioned on X, $H(X|Y)$, is:

$$H(X|Y) = H(X, Y) - H(Y) = -\sum_{i=1}^{N}\sum_{j=1}^{M} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(y_j)}. \tag{3}$$

From these three definitions of entropy, one can define a fourth measure known as 'mutual information,' symbolized by I. The mutual information of two random variables represents how much uncertainty is gained about one variable once the value of the second variable is known. In other words, it measures the amount of overlap or the degree of association between the two variables. Thus, it serves a similar purpose as covariance in variance/covariance analyses (though, see Garner & McGill, 1956, for an analysis of the differences between the two approaches). Formally, the mutual information between X and Y, $I(X; Y)$, is:

$$I(X; Y) = H(X) - H(X|Y) = -\sum_{i=1}^{N}\sum_{j=1}^{M} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)}. \tag{4}$$

There are several useful relationships between the four measures presented above, in addition to $I(X; Y) = H(X) - H(X|Y)$ and $H(X|Y) = H(X, Y) - H(Y)$. They can be found in any textbook on information theory (e.g., Cover & Thomas, 2006).

With the basics of information theory in place, in the next section, I move to Krakauer et al.'s information-theoretic account of individuality.

3 Information-theoretic measures of individuality

Krakauer et al. define several notions of individuality, all of which rely on mutual information. To understand their approach, we must first suppose a discrete Markovian stochastic process defined by a set of random (micro)variables producing data. A stochastic process is a set of random variables indexed over time. A *Markovian* stochastic process is a stochastic process for which the state at a particular time step t depends solely on the state of the process at the time step $t-1$. From there, Krakauer et al. assume that one can coarse-grain these (micro)variables into two (macro)variables: S for 'system' and E for 'environment,' as represented in the directed acyclic graph (DAG) in Fig. 1. Note that the particular way the microvariables are coarse-grained can be purely arbitrary. With n microvariables, there are *a priori* $2^n - 2$ possible ways to partition these into two macrovariables, where S can be composed of one, two, ..., or $n-1$ microvariables, and, consequently, E is composed of $n-1$, $n-2$, ... or one microvariable.[1]

With this setting in place, Krakauer et al. aim to assess whether the macrovariable S, following a particular coarse-graining, scores high in individuality using their information-theoretic measures. Importantly, different coarse-grainings might score identically on individuality. This does not represent a problem for Krakauer et al.'s account, since they embrace the idea that individuals can be nested.

As mentioned in the introduction, at the core of Krakauer et al.'s account is the idea that 'individuals are aggregates that "propagate" information from the past to the future and have temporal integrity' (Krakauer et al., 2020, p. 212). I should clarify here that I understand this statement as the aggregate system itself being *causally* involved in its future state (or that of its offspring). This follows from the minimal conception of what it is to be an individual in the evolutionary or Darwinian sense (Godfrey-Smith, 2009, 2015). I believe this idea can be extended to other conceptions of individuality and that Krakauer et al. would agree with me that a system that would be passively reconstructed through time due to some environmental factors without the capacity to change its future state is not an individual. Rather, it is merely a reconstructed entity.

To formalize their account, Krakauer et al. propose to start from the mutual information of the couple S and E at a particular time t, and S at a later time $t+1$, $I(S_t, E_t; S_{t+1})$. Following the chain rule for mutual information (see Cover & Thomas, 2006, pp. 23–24), this mutual information can be decomposed in two ways:

$$\begin{aligned} I(S_t, E_t; S_{t+1}) &= \overbrace{I(S_{t+1}; S_t)}^{A^*} + \overbrace{I(S_{t+1}; E_t|S_t)}^{nC} \\ &= \underbrace{I(S_{t+1}; E_t)}_{\mathcal{E}i} + \underbrace{I(S_{t+1}; S_t|E_t)}_{A}. \end{aligned} \tag{5}$$

We can see from Eq. 5 that the first way $I(S_t, E_t; S_{t+1})$ can be decomposed is as the sum between $I(S_{t+1}; S_t)$, the mutual information between S at t and at $t+1$, and $I(S_{t+1}; E_t|S_t)$, the mutual information between E at t and S at $t+1$ conditioned on

[1] Note that the same partition would be considered the 'environment' in half the combinations and the 'system' in the other half.

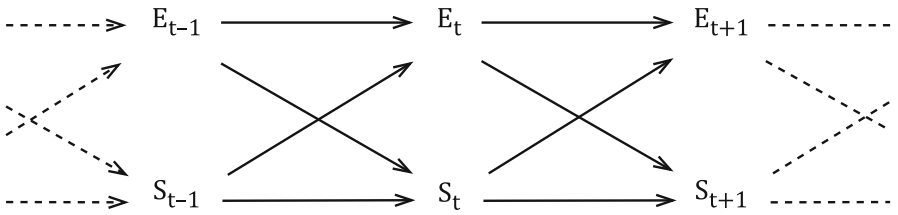

Fig. 1 Directed acyclic graph of the system-environment interaction proposed by Krakauer et al. Modified from Krakauer et al. (2020)

S at t. Following Krakauer et al., I will refer to these two terms as 'A*' and 'nC,' for 'autonomy' and 'non-closure,' respectively. Intuitively, if the state of the system at $t+1$ is entirely explained by the state of the system at t, the environment at t had no influence on the system at $t+1$; thus, nC is nil, or, in other words, the system is informationally closed. Krakauer et al. associate a high value of A* with a type of individuality they call 'organismal individuality,' noting that 'it should be high when the system is largely in control of its environment' (p. 214). In contrast, they associate a high value of nC with a type of individuality they call 'environmental[ly] determined individuality' (p. 215).

The second way to decompose $I(S_t, E_t; S_{t+1})$ from Eq. 5 is as the sum between $I(S_{t+1}; E_t)$, the mutual information between E at t and S at $t+1$, and $I(S_{t+1}; S_t|E_t)$, the mutual information between S at t and at $t+1$ conditioned on E at t. I will refer to these two terms as '$\mathcal{E}i$' for 'environmental influence' and, following Krakauer et al., 'A' for 'autonomy' in a distinct sense from A*, respectively. An analysis of 'A' and 'A*' and their interpretation as forms of autonomy is explored in Bertschinger et al. (2008). Again, intuitively, if the state of the system at $t+1$ is fully explained by the state of the environment at t, the system at t did not influence the system at $t+1$; consequently, A is nil. When A is maximal, the state of the system at t does not depend on the state of the environment at $t-1$. Krakauer et al. associate this property with a type of individuality they call 'colonial individuality' and propose microbes as an example because they 'share only a small amount of information with the environment in which they live' (p. 215).[2]

In the next section, I argue that Krakauer et al.'s proposal that these three measures refer to different concepts (or types) of individuality is in tension with the idea that an (informational) individual has temporal integrity and propagates information through time, where propagation is interpreted causally.

4 Against pluralism about individuality

In the previous section, we saw that starting from the two decompositions in Eq. 5, Krakauer et al. proposed that three of the four terms—namely A*, A, and nC—represent concepts or types of individuality. An alternative way to consider Krakauer

[2] Krakauer et al. propose more fine-grained decompositions of the four measures—A*, A, nC, and $\mathcal{E}i$—following Williams and Beer (2010). However, they note that Williams and Beer's proposal has been criticized and no consensus has been reached on an alternative decomposition. For this reason, I will not present these decompositions here.

et al.'s approach would be as providing different measures that refer to a single concept of individuality. nC, A^*, and A would each capture partially or indirectly what individuality entails. Following this interpretation, Krakauer et al. would be 'measure pluralists' but 'concept monists' about individuality. While this is a possibility, and Krakauer et al. do not dismiss it outright, I do not think this corresponds to their view considering, as we saw in the previous section, that they explicitly associate each measure with a different type of individuality (p. 215). Further, they do not provide any reason to favor one of the three measures.

If my reasoning is correct, it is only fair to consider Krakauer et al. as concept pluralists regarding individuality. In principle, I have nothing against this type of pluralism. Different concepts of individuality can play different roles in different contexts (see Clarke, 2010; Lidgard & Nyhart, 2017; Wilson & Barker, 2019). However, if the different concepts are supposed to refer to the very same idea—the view that an individual is an entity that propagates information through time and has some integrity—there is ground to explore whether one of the three measures is better than the others at characterizing individuality in that respect.

Krakauer et al. remain silent on this latter point. However, they should not be blamed for this, as I see their contribution as foundational—that is, infusing the literature with new ways of thinking about individuality. That being said, the question of the possible superiority of one of the three measures remains. This section addresses this issue. I argue that the measures nC and A^*, proposed by Krakauer et al., do not correspond to an adequate *concept* of individuality as a system propagating information through time with some integrity. In the following section, I show that, in the particular setting proposed by Krakauer et al., A performs better than the two others. However, it remains an associative rather than a causative measure. From there, I propose my own information-theoretic measure of individuality that better corresponds to Krakauer et al.'s verbal formulation.

Starting with nC, the reason it should be discarded as a measure directly characterizing individuality is rather simple. nC measures a relationship between E and S over time, while individuality solely concerns S over time, following Krakauer et al.'s own proposal. However, it should be clear that rejecting nC as a measure corresponding to the proposal that an individual is a system propagating information does not imply that this measure should be regarded as generally useless. On the contrary, as we saw, nC provides a measure of the system's sensitivity to environmental influences. If nC is small, environmental influence is small; thus, the integrity of the system is retained over time. However, this does not, in and of itself, give us an indication of the system's degree of individuality. This is so because the system might also be a very poor predictor of its state in the future. In consequence, a low nC could be measured in both systems that would not be considered by anyone as individuals *and* systems that are indubitably individuals. Assuming the same degree of individuality between the two systems, a different value of nC between these two systems might help us evaluate the degree to which these systems differ in their intrinsic capacity to change under new environmental conditions.[3] This feature might be an important

[3] This capacity is sometimes called 'evolvability' (see Brown, 2014 for an analysis).

property of individuality if the notion of individuality is understood more broadly than informational individuality, as proposed by Krakauer et al.

Moving to A^* and A, the main issue with these measures is that they are associative rather than causal. This is so because mutual information (like covariance) is a measure of dependence, not causation, between two variables. To be clear, this problem is also encountered with nC, but we already established that it does not represent an adequate measure of individuality *per se*.

To understand this problem, let us start with A^* (A will be discussed in the next section) and notice that high mutual information between S at t and at $t+1$ does not mean that a prior state of S is *causally* responsible for this association. This is so because E at time t might be a confounding variable. Similarly, a high covariance (or correlation, which is a standardized covariance) between two variables does not mean that one causes the other. Yet, as I argued above, I take it that in the statement that an individual S propagates information, earlier states of S must be a cause of its subsequent states, not merely that the states are simply correlated. This represents a problem because we could imagine a setting in which the environment at any time step fully determines both the system's and environment's state at the next time step, as represented in the DAG in Fig. 2. In such a case, S_t and S_{t+1} would be associated, but the relationship would not be causal, only correlational. In the next section, starting from the measure proposed by Krakauer et al., I provide a *causal* information-theoretic measure of individuality that lives up to the causal requirement.

5 A causal information-theoretic measure of individuality

The interventionist account of causation (Woodward, 2003, 2016; Pearl, 2009; Pearl et al., 2016) has been developed to distinguish whether two variables are causally related rather than merely associated. Following this account, X is a cause of Y if intervening on the value of X leads to a change in the value of Y. An intervention on a variable is defined as a change in the value of this variable that does not lead to any other change in any other variable that is also causally upstream of the putative effect variable of interest at the time the intervention is performed. Formally, an intervention is characterized with the $do(.)$ operator, symbolized here with a '$\widehat{\ }$.'

Importantly, it should be noted that an intervention is an idealization; as such, it can never be observed in the real world. Randomized control trials (RCTs) and some

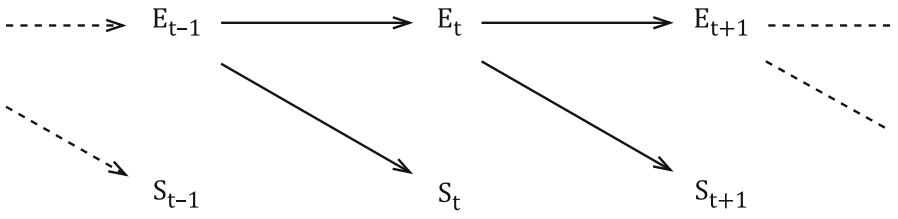

Fig. 2 Directed acyclic graph in which the environment fully determines the state of the system and its state at the next time step. This creates a correlation between the state of the system at a given time step and that of the next one

very controlled experiments (e.g., in physics) are the most faithful way to emulate an ideal intervention in the physical world. In the case of an RCT, each participant of the trial is assigned randomly one of the different treatment groups (including a control one). This procedure ensures that if the number of participants is large enough, any difference in outcome between these groups will be due to the difference in treatment. Randomization ensures that any variable that would otherwise be correlated with the treatment variable (i.e., a confounding variable) is decorrelated from the treatment variable. Thus, any difference observed between the different groups can be interpreted as if it were the result of an ideal intervention.

When dealing with observational data, by definition, no experimental procedure can eliminate confounding variables. Nonetheless, assuming a particular causal model represented by a DAG, an ideal intervention can be emulated if the paths between the variables of the candidate causal relationship satisfy a number of properties. If these properties—the exposition of which would go well beyond the scope of this paper—are satisfied, an ideal intervention can be emulated using an adjustment formula (Pearl et al., 2016, chap. 3). This formula allows us to express the outcome of the $do(.)$ operator on a variable using only standard conditional probabilities.

Recently, Griffiths et al. (2015; see also Korb et al., 2011; Ay & Polani, 2008; Pocheville et al., 2017) have proposed information-theoretic measures of causal relationships based on mutual information within the interventionist framework. I will apply them here in the context of individuality. Starting from the definition of an individual as an aggregate that (causally) propagates information through time, within Krakauer et al.'s framework, we can define a measure of individuality as the causal mutual information between the system at t and $t+1$, so that:

$$\widehat{A} = I(S_{t+1}; \widehat{S_t}), \tag{6}$$

where $\widehat{S_t}$ represents the variable S intervened upon at t.[4]

Unlike A^*, which is a measure of *association* between S_t and S_{t+1}, $\widehat{A}$ represents a *causal* measure of individuality that better corresponds to the idea that an individual *causally propagates* information. If we take again a setting in which the environment at t fully determines the system at $t+1$, one would find that $\widehat{A}$ is 0 because there would be no causal relation between S_t and S_{t+1}, while A^* might be positive simply because the environment is a common cause of the system over two or more successive time steps as would be the case in Fig. 2.

Until now, I have argued that $\widehat{A}$ represents a better measure of individuality than A^*. This is so because it is a *causal* rather than an *associative* measure. However, I have said nothing about A. As mentioned earlier, an intervention is an *idealization*. Recall that this means it cannot be enacted in the real world but only emulated following an adjustment formula that involves conditioning on factors that could be potential confounders (see Pearl et al., 2016, chap. 3). Once this adjustment formula is applied to the model proposed by Krakauer et al., the minimal conditioning that yields an equiva-

[4] Note that this and other causal measures rely on a probability distribution for the intervened upon states of $\widehat{S}$. It might be chosen to be uniform, identical to the distribution of S in the population studied, or again within a 'normal' range of values, depending on the explanatory goals of the measure. See Griffiths et al. (2015) and Pocheville et al. (2017) for discussions of this point.

lence with $\widehat{A}$ from observations is E_t, so that $\widehat{A} = I(S_{t+1}, \widehat{S_t}) \equiv I(S_{t+1}; S_t|E_t) = A$. Thus, in this setting, A corresponds to $\widehat{A}$. However, it is important to note that $\widehat{A}$ and A are not equal in general. This is so because they have been derived from two different perspectives, only one of which explicitly employs causal reasoning. Following a causal conception of individuality, $\widehat{A}$ corresponds better to the idea that a system causally propagates information.

To see the importance of this point, we need to move away from Krakauer et al.'s specific model. Suppose a setting in which the time does not represent a point but a slice (T), in such a way that the system and the environment can influence each other not only between time slices but also within time slices. We further assume that the environment is fine-grained into two variables, E_a and E_b, and that the following relationships between variables hold: E_a influences S within time slices and influences its state between time slices; E_b influences both its state and the state of S between time slices; finally, S influences E_b within time slices and its state between time slices. The DAG of this setting is presented in Fig. 3. It is a simple example of a (dynamic) 2-time-slice Bayesian network (see Koller, 2009, chap. 6, for details).

In this model, $\widehat{A}$ would still be defined as $I(S_{T+1}, \widehat{S_T})$. However, the adjustment to obtain an equivalent measure to $\widehat{A}$ would not be $I(S_{T+1}; S_T|E_T) = I(S_{T+1}; S_t|E_{a_T}, Eb_{b_T}) = A$. This is because adjusting for E_{b_T} here would eliminate the indirect causal effect of S_T on S_{T+1} mediated by the Eb_T. Instead, the correct adjusted measure would be $I(S_{T+1}; S_T|E_{a_T})$.

Before moving on, I should note that some of the authors in Krakauer et al. (2020) developed causal notions of informational flow and autonomy in Ay and Polani (2008) and Bertschinger et al. (2008), respectively, which are similar to the measure $\widehat{A}$ proposed here. In particular, Bertschinger et al. (2008) showed the limits of a correlative notion of autonomy. Given the explicit links between autonomy and informational individuality, it is surprising that Krakauer et al. did not extend their work to take into account the problems of non-causal association.

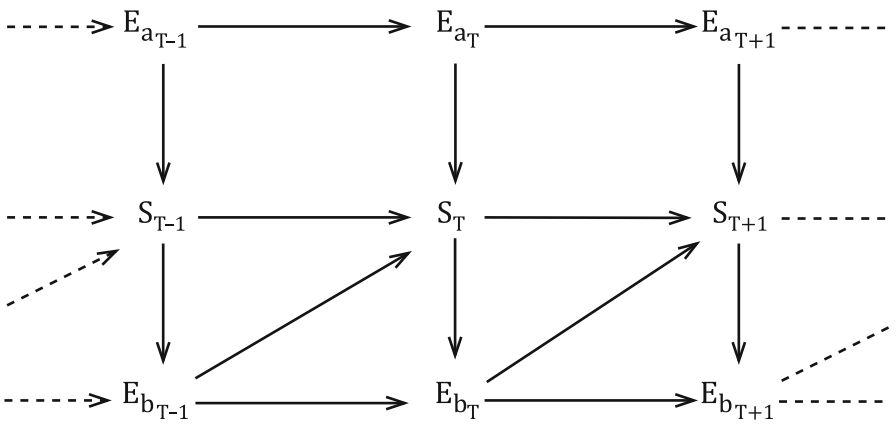

Fig. 3 Directed acyclic graph of a (dynamic) 2-time-slice Bayesian network in which the system and environment interact. The environment is fine-grained into two variables, E_a and E_b. T represents a time slice rather than a point in time

Having proposed an alternative conception of informational individuality to the one proposed by Krakauer et al., in the next section, I turn to ETIs and show how this account of individuality can be deployed and elaborated in this context.

6 Evolutionary transitions in individuality and causal information

Modern organisms are the result of a succession of ETIs that have occurred numerous times during the history of life (Maynard-Smith & Szathmary, 1995; Boomsma, 2022; Bourke, 2011; Michod, 1999; Black et al., 2020; Calcott & Sterelny, 2011; Bouchard & Huneman, 2013). Examples of ETIs include the transitions from molecules to primordial cells, from unicellular to multicellular organisms, and from multicellular organisms to super-organisms such as ant colonies or bee hives (see Bourke, 2011 for a review of the different types of ETIs with examples). More abstractly, an ETI occurs when individuals at a given level of organization (hereafter, 'particles') start to interact in such a way as to produce higher-level entities (hereafter, 'collectives') that are subsequently recognized as individuals in their own right and define a new level of organization.

A popular approach to define a new collective level of individuality, including as a result of an ETI, is from the perspective of Lewontin's (1970, 1985) three conditions for evolution by natural selection: variation, fitness differences, and heritability at that level. However, as was argued by Griesemer (2000), the three conditions presuppose the existence of units. The project of providing an account of an ETI is precisely to find the conditions of the emergence of new units of selection or new levels of individuality.[5] Therefore, while the three conditions might be regarded as necessary, or at least important, for evolution by natural selection to be possible at any level of individuality,[6] they are, in and of themselves, insufficient to define a new level of individuality. At the very least, they must be complemented with another approach that points to some properties of individuality.

Although Krakauer et al. do not discuss how their approach could complement Lewontin's approach to Darwinian individuality, they connect it to the multilevel version of the Price equation (p. 220). The Price equation is a tool of choice in evolutionary theory (see Luque, 2017; Okasha, 2006; Price, 1970, 1972; Rice, 2004). Okasha (2006) provides an analysis of which the conclusion is 'that Price's equation *almost* vindicates the Lewontin conditions' (p. 37). In Bourrat (2021a), I further demonstrated that some of the limitations classically associated with the Price equation also apply to Lewontin's conditions given the similarities between the two approaches. In particular, to apply the multilevel version of the Price equation, one must partition a population of particles into collectives. However, it is well known that the specific partitioning used can be purely arbitrary (Okasha, 2006; Heisler & Damuth, 1987; Damuth, 1985; Bourrat, 2021a, d). Thus, no justification for the particular partitioning used in the mul-

[5] I will consider here that a new level of individuality is equivalent to a unit of selection.

[6] Godfrey-Smith (2007) provides an analysis in which he shows that the conditions are not sufficient for evolution by natural selection.

tilevel Price equation can come from the equation itself. Krakauer et al.'s approach (and its causal counterpart proposed here), by providing the resources to establish which partitioning(s) can be considered a unit or informational individual, could be used as a principled way to choose the partitionings that characterize levels of individuality, thereby usefully complementing a Pricean or Lewontinian approach in an evolutionary context.

Being able to find the relevant units upon which to apply Lewontin's conditions or the multilevel version of the Price equation—an important aspect of ETIs—is not the only use of Krakauer et al.'s framework in the context of ETIs. In the remainder of this section, I show how it can also be deployed and elaborated in the context of the ecological scaffolding model for ETIs. Although I do not show it here, the potential relevance of Krakauer et al.'s framework goes beyond the ecological scaffolding model in the context of ETIs.

To begin, suppose that, in the process of coarse-graining a set of microvariables into two macrovariables, an observer finds a coarse-graining where S_1 and E_1 are the macrovariables (as represented in Fig. 4), where S_1 scores high on $\widehat{A_{S_1}}$. Following the reasoning developed in the preceding sections, this would mean that S_1 represents a *bona fide* characteristic level of informational individuality. Further, assume that there is no other way to coarse-grain the microvariables that leads to the same or a higher score for $\widehat{A}$. We now let the system evolve for some time and want to know whether an ETI is occurring. How shall we proceed?

Ecological scaffolding has recently been proposed as a mechanism by which a transition could occur (Black et al., 2020). According to this model, an ecological scaffold—that is, specific ecological factors—can initiate the formation of collectives exhibiting Lewontin's three conditions. Initially, the three conditions are better characterized as Lewontin-*like*, following Bourrat's (2022) terminology, because if the ecological scaffold were to be removed, the collectives would disintegrate. Progressively, however, as the transition proceeds, the scaffold becomes endogenized so that even once the ecological conditions change or if the scaffold were removed, the integrity of collectives would remain. Under one possible scenario of ecological scaffolding, the scaffold results from the particles' activities, similar to niche construction (see Bourrat, 2022, for discussion). Black et al. (2020) provide several biological situations in which the ecological scaffolding model could be relevant. Note, importantly, that the ecological scaffolding model and, more particularly, the scenario involving niche construction is by no means the only path through which an ETI can occur (Bourrat et al., 2023). I only use this scenario here as an illustration.

Starting from the causal information-theoretic framework proposed here, an ETI being initiated through the model just outlined would result in the quantity $I(E_{1_{t+1}}; \widehat{S_{1_t}})$ increasing over time, implying S_1 has more (causal) control over the environment through time. Note that neither this quantity nor its associative counterpart $I(E_{1_{t+1}}; S_{1_t})$ is discussed by Krakauer et al. Second, following a new coarse-graining of the microvariables, where S_2 and E_2 are the macrovariables, one in which S_2 would comprise S_1 and part of E_1, as illustrated in Fig. 4b, $\widehat{A_{S_2}}$ would also increase over time. One possible criterion **C1** that could be used to establish that the transition (or at least that the scaffolding process) has occurred is the following:

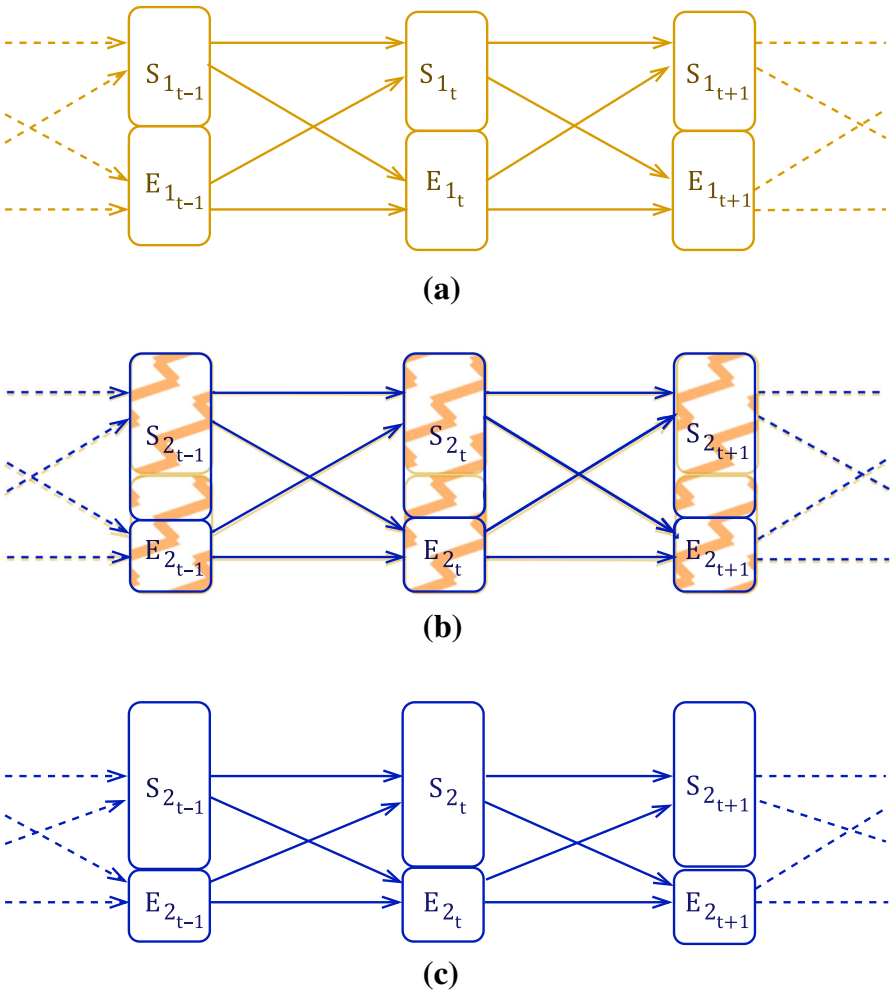

Fig. 4 Directed acyclic graphs illustrating the ecological scaffolding model for evolutionary transitions in individuality (ETIs). (a) We start with the coarse-graining that separates a system and an environment into S_1 and E_1, respectively, where the S_1 is an individual. We let the population evolve for some time. (b) At some point, a second coarse-graining separating a system and an environment into S_2 and E_2 and where S_2 contains some of the microvariable initially present in E_1 permits defining a system that propagates information through time. (c) This new coarse-graining subsequently leads to a level of individuality comparable to that of S_1, at which point an ETI has occurred

C1 *Assuming that S_1 is indubitably a level of individuality, S_2 represents a potential level of individuality if*

$$\frac{\widehat{A_{S_2}}}{H(\widehat{S_{2_t}})} \geq \frac{\widehat{A_{S_1}}}{H(\widehat{S_{1_t}})}\bigg|_{\theta_0}.$$

This requires unpacking. First, the rationale behind the criterion is that if S_1 characterizes a level of individuality, its level of individuality can be used as a reference

point to assess the degree of individuality of S_2. From there, we can consider that S_2 defines a new level of individuality if the measure of $\widehat{A_{S_2}}$ is at least as high as $\widehat{A_{S_1}}|_{\theta_0}$, as illustrated in Fig. 4c: where, at time θ_0, an ETI has not occurred.

Second, there is an important reason why $\widehat{A_{S_1}}|_{\theta_0}$ and $\widehat{A_{S_2}}$ must be compared while normalized with $H(\widehat{S_{1_t}})|_{\theta_0}$ and $H(\widehat{S_{2_t}})$, respectively, rather than using the absolute values. Assuming that the total number of values for the microvariables remains the same throughout the ETI, if S_2 includes part of E_1, this means that $\widehat{A_{S_2}}$ could be higher than $\widehat{A_S}$, not because S_2 has a higher level of individuality but because it exhibits a higher number of possible states: that is, $H(\widehat{S_{2_t}}) > H(\widehat{S_{1_t}})|_{\theta_0}$. Using normalized measures of $\widehat{A}$ eliminates this potential reason for the difference observed.

One drawback of normalization is that a system with very few possible states might be regarded as incompatible with scoring high on individuality. Using only normalized versions of $\widehat{A}$ would not allow discriminating systems exhibiting a low number of possible states from those exhibiting a high number of states. One way to mitigate this problem would be to use the normalized version of $\widehat{A}$ only for systems with a $H(\widehat{S_t})$ above a certain absolute threshold, to be determined by the observer, compatible with individuality.

Establishing that the coarse-graining using S_2 and E_2 as macrovariables satisfies **C1** would nevertheless be insufficient to show that an ETI has occurred. **C1** is only a criterion for *potential* individuality. This is so because, following the ecological scaffolding model, for an ETI to be complete, collectives must retain their integrity once the scaffold is lifted—that is, the scaffold must be endogenized (Black et al., 2020; Bourrat, 2022; Doulcier et al., 2023; Bourrat et al., 2023; Veit, 2021; Neto et al., 2023). Recall that, according to the ecological scaffolding scenario for ETIs, when the scaffold is present, the units of the population are only Darwinian-like. One way to assess whether the collectives defined by the coarse-graining where S_2 and E_2 are the macrovariables have endogenized their scaffold, and are thus truly Darwinian, would be to measure whether their capacity to propagate information is sensitive to environmental changes. If it is, this would be evidence that they have not endogenized their scaffold and, thus, are not higher-level individuals. On the contrary, if the information is propagated even when the environment changes, this would be evidence that their capacity to propagate information in time does not depend on some particular configuration of the environment (i.e., a scaffold).

One way to implement this second criterion **C2**, thereby extending Krakauer et al.'s proposal, would be by measuring the variance of $\widehat{A_{S_2}}$ under intervention on E_2 so that:

C2 *S_2 defines a new level of individuality if the variance of $\widehat{A_{S_2}}$ under the possible interventions on E_{2_t} is low—that is, (eliminating the subscript t for clarity)* $\mathrm{Var}(\widehat{A_{S_2}}|\widehat{E_2} = \widehat{e_{2_k}})$, *where e_{2_k} is a possible state of E_2 at t, is low.*

This condition of invariance is inspired by the analysis of causal invariance and stability initially proposed by Woodward (2000) and for which information-theoretic measures have been proposed (see Bourrat, 2021c). Note that a measure similar to $\widehat{A_{S_2}}|\widehat{E_2} = \widehat{e_{2_k}}$ has been proposed as a measure of information flow in a system by Ay and Polani (2008).

7 Conclusion and future directions

In this paper, I made some connections between a new proposal for characterizing the concept of individuality in terms of information theory and the well-known interventionist account in the philosophy of causation. Further, I showed how this proposal could be deployed and elaborated to characterize an ETI following the ecological scaffolding model.

The framework proposed by Krakauer et al. is highly abstract; as such, it has great potential to be applied across different scales and different domains. However, the framework is not without limitations. For instance, even in its causal version, the measure of individuality proposed by Krakauer et al. cannot, in and of itself, differentiate a genuine level of individuality from a coarse-graining that would characterize only part of a level of individuality, such as tissues in multicellular organisms or half-organisms. I should mention that one prima facie way to fix this problem would be to consider that, assuming a particular timescale, individuals are only found at the coarsest grain of description in which $\widehat{A}$ is maximal. Future work should make this condition more precise and amend $\widehat{A}$ accordingly.

Another area of exploration worth mentioning is the links between $\widehat{A}$ and heritability. Considering that heritability is a staple of evolutionary theory (one of Lewontin's conditions) and relies on a decomposition between genotype and environment that is similar to Krakauer et al.'s system/environment decomposition, it would be worthwhile to explore the connection between the two measures and their respective frameworks to determine whether they face the same limitations. Some work has been initiated in this area by drawing some links between heritability and mutual information (see Bourrat, 2021b). We also know that the Price equation mentioned above can be derived in a form that makes apparent a term of heritability (Rice, 2004; Okasha, 2006; Queller, 1992; Bourrat, 2015). Frank (2012) initiated work to provide a link between information theory and the Price equation. It would be worthwhile to assess whether the heritability term in the Price equation can be connected to the notion of informational individuality proposed here.

Acknowledgements I am grateful to two anonymous reviewers for their comments on previous versions of the manuscript. I thank Guilhem Doulcier, Katrin Hammerschmidt, and Peter Takacs, for discussions on the topic of biological individuality, as well as the audience at PSA 2022 where a preliminary version of this work was presented. The author gratefully acknowledges the financial support of the John Templeton Foundation (#62220). The opinions expressed in this paper are those of the author and not those of the John Templeton Foundation. This research was also supported under the Australian Research Council's Discovery Projects funding scheme (Project Number DE210100303).

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Declarations

Conflicts of interest The author has no financial or non-financial interests that are directly or indirectly related to the work submitted for publication.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give

appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Ay, N., & Polani, D. (2008). Information flows in causal networks. *Advances in Complex Systems, 11*(1), 17–41.

Bertschinger, N., et al. (2008). Autonomy: An information theoretic perspective. *Biosystems, 91*(2), 331–345.

Black, A. J., Pierrick, B., & Rainey, P. B. (2020). Ecological scaffolding and the evolution of individuality. *Nature Ecology & Evolution, 4*, 426–436.

Boomsma, J. J. (2022). *Domains and major transitions of social evolution*. Oxford University Press.

Bouchard, F., & Huneman, P. (2013). *From groups to individuals: Evolution and emerging individuality*. MIT Press.

Bourke, A. F. G. (2011). *Principles of social evolution*. Oxford University Press.

Bourrat, P., Takacs, P., Doulcier, G., Nitschke, M., Black, A., Hammerschmidt, K., & Rainey, P. (2023). *Individuality through ecology*. https://doi.org/10.31219/osf.io/9h26t

Bourrat, P. (2015). How to read 'heritability' in the recipe approach to natural selection. *The British Journal for the Philosophy of Science, 66*(4), 883–903.

Bourrat, P. (2019). Variation of information as a measure of one-to-one causal specificity. *European Journal for Philosophy of Science, 9*(1), 11.

Bourrat, P. (2021a). *Facts, conventions, and the levels of selection. Elements in the Philosophy Biology*. Cambridge University Press.

Bourrat, P. (2021c). Measuring causal invariance formally. *Entropy, 23*(6), 690.

Bourrat, P. (2021d). Transitions in evolution: A formal analysis. *Synthese, 198*(4), 3699–3731.

Bourrat, P. (2022). Evolutionary transitions in individuality by endogenization of scaffolded properties. *The British Journal for the Philosophy of Science*. https://doi.org/10.1086/719118

Bourrat, P. (2023). A coarse-graining account of individuality: How the emergence of individuals represents a summary of lower-level evolutionary processes. *Biology & Philosophy, 38*(4), 33.

Bourrat, P. (2021). Heritability, causal influence and locality. *Synthese, 198*(7), 6689–6715.

Brown, R. L. (2014). What evolvability really is. *The British Journal for the Philosophy of Science, 65*(3), 549–572.

Calcott, B., & Sterelny, K. (2011). *The major transitions in evolution revisited*. MIT Press.

Clarke, E. (2010). The problem of biological individuality. *Biological Theory, 5*(4), 312–325.

Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory*. John Wiley & Sons.

Damuth, J. (1985). Selection among "species": A formulation in terms of natural functional units. *Evolution, 39*(5), 1132–1146.

Doulcier, G., Takacs, P., Hammerschmidt, K., & Bourrat, P. (2023). Stability of ecologically scaffolded traits during evolutionary transitions in individuality. *bioRxiv*, 2023.08.17.553478. https://doi.org/10.1101/2023.08.17.553478.

Frank, S. A. (2012). Natural selection. III. Selection versus transmission and the levels of selection. *Journal of Evolutionary Biology, 25*, 227–243.

Garner, W. R., & McGill, W. J. (1956). The relation between information and variance analyses. *Psychometrika, 21*(3), 219–228.

Godfrey-Smith, P. (2007). Conditions for evolution by natural selection. *Journal of Philosophy, 104*(10), 489–516.

Godfrey-Smith, P. (2009). *Darwinian populations and natural selection*. Oxford University Press.

Godfrey-Smith, P. (2015). Reproduction, symbiosis, and the eukaryotic cell. *Proceedings of the National Academy of Sciences, 112*(33), 10120–10125.

Griesemer, J. R. (2000). The units of evolutionary transition. *Selection, 1*(1-3), 67–80.

Griffiths, P. E. et al. (2015). Measuring causal specificity. *Philosophy of Science, 82*(4), 529–555.

Heisler, I. L., & Damuth, J. (1987). A method for analyzing selection in hierarchically structured populations. *The American Naturalist, 130*(4), 582–602.
Koller, D. (2009). *Probabilistic graphical models: Principles and techniques* (1st ed.). MIT Press Academic.
Korb, K. B., Nyberg, E. P., & Hope, L. (2011). A new causal power theory. In P. M. Illari, F. Russo, & J. Williamson (Eds.), *Causality in the sciences* (pp. 628–652). Oxford University Press.
Krakauer, D. et al. (2020). The information theory of individuality. *Theory in Biosciences, 139*(2), 209–223.
Lewontin, R. C. (1970). The units of selection. *Annual Review of Ecology and Systematics, 1*(1), 1–18.
Lewontin, R. C. (1985). Adaptation. In R. Levins, & R. C. Lewontin (Eds.), *Dialectics and reductionism in ecology* (pp. 65–84). Harvard University Press.
Lidgard, S., & Nyhart, L. K. (2017). The work of biological individuality: Concepts and contexts. In S. Lidgard, & L. K. Nyhart (Eds.), *Biological individuality: Integrating scientific, philosophical, and historical perspectives* (pp. 17–62). University of Chicago Press.
Luque, V. J. (2017). One equation to rule them all: A philosophical analysis of the price equation. *Biology & Philosophy, 32*(1), 97–125.
Maynard-Smith, J., & Szathmary, E. (1995). *The major transitions in evolution*. Oxford Univeristy Press.
Michod, R. E. (1999). *Darwinian dynamics*. Princeton University Press.
Neto, C., Meynell, L., & Jones, C. T. (2023). Scaffolds and scaffolding: An explanatory strategy in evolutionary biology. *Biology & Philosophy, 38*(2), 8.
Okasha, S. (2006). *Evolution and the levels of selection*. Oxford University Press.
Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal inference in statistics: A primer*. John Wiley & Sons.
Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge University Press.
Pocheville, A., Griffiths, P. E., & Stotz, K. (2017). Comparing causes – an information-theoretic approach to specificity, proportionality and stability. In H. Leitgeb, I. Niiniluoto, E. Sober, & P. Seppälä (Eds.), *Proceedings of the 15th Congress of Logic, Methodology and Philosophy of Science* (pp. 260–275). College Publications.
Price, G. R. (1970). Selection and covariance. *Nature, 227*(5257), 520–21.
Price, G. R. (1972). Extension of covariance selection mathematics. *Annals of Human Genetics, 35*, 485–490.
Queller, D. C. (1992). Quantitative genetics, inclusive fitness, and group selection. *The American Naturalist, 139*(3), 540–558.
Rice, S. H. (2004). *Evolutionary theory: Mathematical and conceptual foundations*. Sinauer Associates.
Shannon, C. (1948). A mathematical theory of communication. *The Bell System Technical Journal, 27*, 379–423.
Veit, W. (2021). Scaffolding natural selection. *Biological Theory*. https://doi.org/10.1007/s13752-021-00387-6
Williams, P. L., & Beer, R. D. (2010). *Nonnegative decomposition of multivariate information*. https://doi.org/10.48550/arXiv.1004.2515
Wilson, R. A., & Barker, M. (2019). Biological individuals. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.
Woodward, J. (2000). Explanation and invariance in the special sciences. *The British Journal for the Philosophy of Science, 51*(2), 197–254.
Woodward, J. (2016). Causation and manipulability. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.
Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford University Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.