

Information channels and biomarkers of disease

Phyllis Illari and Federica Russo

Draft of November 10, 2014

Abstract

Current research in molecular epidemiology uses biomarkers to model the different disease phases from environmental exposure, to early clinical changes, to development of disease. The hope is to get a better understanding of the causal impact of a number of pollutants and chemicals on several diseases, including cancer and allergies. In a recent paper Russo and Williamson (2012) address the question of what evidential elements enter the conceptualisation and modelling stages of this type of biomarkers research. Recent research in causality has examined Ned Hall's distinction between two concepts of causality: production and dependence (Hall, 2004). In another recent paper, Illari (2011b) examined the relatively under-explored production approach to causality, arguing that at least one job of an account of causal production is to illuminate our inferential practices concerning causal linking. Illari argued that an informational account solves existing problems with traditional accounts. This paper follows up this previous work by investigating the nature of the causal links established in biomarkers research. We argue that traditional accounts of productive causality are unable to provide a sensible account of the nature of the causal link in biomarkers research, while an informational account is very promising.

1 Introduction – What is the problem?

Epidemiology has long established that environmental factors are important causes of disease, but research so far has established only that broad classes of environmental factors are correlated, albeit robustly, to classes of diseases. Current research in molecular epidemiology aims to improve our understanding of causal links between environmental factors and disease. In order to do that scientists carry out precise measurements of levels of specific chemicals, and measurements of changes that happen in the body at the molecular level, in response to exposure to such chemicals. This is the core idea behind biomarkers research, or, as the project we will discuss is called, 'EXPOsOMICS' (<http://www.EXPOsOMICSPROJECT.eu>). What is sought is more precise assessment of environmental exposure on the one hand, and identification of key stages in the development or evolution of disease on the other hand.

This is frontier research, in which the European Union is investing a lot of money, in the hope of making a significant move beyond the results of genomics research, which were disappointing—at least in comparison with the overblown expectations. It is exciting and challenging research, using methods that are

new in their field. It is also highly interdisciplinary, involving at the least biology, bioinformatics, sophisticated statistics, etc. The novelty of the research goes beyond interdisciplinarity, though, as it requires rethinking the conceptual toolbox currently available for thinking about causality. In this paper, we will be interested in whether we have good philosophical concepts to ask and answer methodological, epistemological, and metaphysical questions raised by this kind of research.

In pursuing this aim, this paper follows up earlier attempts to design concepts that help philosophy of causality and scientific practice make progress in saying what causality is, how we reason about it, and what it is that we track when we track causal relations.

Russo and Williamson (2012) explored the epistemology underlying biomarkers research, arguing that scientists need evidence of difference-making and of mechanisms to establish causal claims. The way scientists model the progress—or, as they say, ‘evolution’—of biomarkers from exposure to disease borrows elements of processes and elements of complex-systems. The integration of evidence of difference-making and of mechanisms gives a better understanding of the epistemology of difficult cases such as biomarkers. Nevertheless, the account offered in that paper was still too coarse-grained to offer a concept of causal linking from macro to micro factors.

Illari (2011b), on the other hand, explored the concept of ‘productive’ causality, to use the terminology of Hall (2004) (see discussion in section 3), exploring what the task of an account of causal production is. The paper argued that at least one job of an account of causal production is to give an account of causal linking that can assist our reasoning in ruling in and ruling out possible causes. The paper suggests that an informational approach might be better than available approaches, but this idea needs development and application to a case.

This paper is organised as follows. First, in section 2, we present biomarkers research, drawing out the novel features of the scientific project, and showing how the project strains concepts of causal linking currently available in the philosophical literature on causality. Biomarkers research is an excellent test case for developing a philosophical account of causal production or linking, as it shows the full complexity of science in practice, engaging with a complex system, and looking for causes where they are exceptionally difficult to find. Throughout the paper, we will examine whether we can get a sense of causal linking in such a difficult case. Second, in section 3, we review the major philosophical accounts that contribute to understanding causal linking, and show that they are unable to satisfy the needs of biomarkers research. In section 4, we show what an informational account can do better. In brief, it can do better because it indicates a radically different way forward for addressing what causal linking is, in a way that can illuminate our causal inferential practices. We finish, in section 5, by showing how an informational account can ally with useful elements of previous accounts, particularly a mechanistic approach.

2 Tracing the evolution of biomarkers

2.1 Environmental exposure and the ‘exposome’

There are two previous approaches to finding causes of disease that are relevant to the creation of the new approach of EXPOsOMICS, using biomarkers of disease. The first is traditional epidemiology. It is known that exposure to environmental factors cause certain diseases. These macro-environmental factors are usually a hodgepodge of various chemicals in the air we breathe, in the water we drink or swim in, in the food we eat, in the products we use for cleaning or cosmetics, etc. The effects observed are also ‘macro’: the development of diseases such as cancers, allergies, etc. Traditional epidemiology can find out what kind of people are most exposed to environmental factors, in terms of some specific health characteristics, or socio-economic status, or occupation. A common approach is to assess exposure in a ‘coarse-grained’ way through questionnaires and interviews, and construct macro-categories of exposure. These approaches provide very useful information about a possible mapping between these macro-categories of exposure and disease. Yet this is all at the ‘macro’ level. Further, research tends to be organised in groupings, studying, for example, risks of water pollution, or risks associated with different occupations, and so is likely to miss interactions between these broad classes of factors. As a result, a great deal is still unknown about long-term low-dose exposure to environmental factors, which are particularly likely to be significant factors in chronic diseases.

The second previous approach is genomics. Researchers sought to explain health and disease through genetic features, changes, and modifications. Multiple technologies were developed, including: genomics to measure what is happening to genes, transcriptomics to measure what genes are being transcribed, proteomics which measures what proteins are being made in the cell, and metabolomics which measures the levels of particular metabolites and so various activities in the cell. In short, omics allow us to track a great deal of what happens in the cell, from DNA, through transcription, to proteins made, to metabolic processes. These technologies are often collectively referred to as the ‘omics’ technologies, for obvious reasons. The mapping of the human genome had enormously high stakes. Despite the massive quantities of data collected, relatively little has been explained about disease causation by genes. Specifically, mapping the genome, and genome-wide association studies (GWAS), indicate that the genome cannot be the whole story. There is strong gene-gene interaction in causing biological features, including disease. We found many gene-disease relationships, but only a few that were as useful as had been hoped. In addition to this, there are important gaps in the story about disease causation due to genetic factors.

Biomarkers research is an emergent field that uses new technologies, particularly omics technologies, to improve the assessment of exposure to environmental factors. A biomarker is a ‘biological marker’. It is a characteristic which is objectively measured and that indicates normal biological processes, or pathogenic processes, or a pharmacological reaction to a therapeutic intervention. The idea is to measure exposure from biological samples, such as of blood and other bodily fluids, to yield a much finer grained analysis of exposure. Biomarkers of some kinds of exposure, and of some kinds of diseases, have been found and

validated, but much remains to be done. Working with biomarkers allows a great deal more specificity in assessing environmental causes of disease.

Finally, the concept of the ‘exposome’ arises. The exposome is the totality of both external exposure to environmental factors, and also internal exposure, which is exposure as measured within the body. The body has an internal chemical environment, and the internal exposure includes biomarkers of exposure to environmental factors, but also chemicals the body itself produces as a result of, for example, inflammation due to environmental factors—see Wild (2005, 2009, 2011); Rappaport (2011); Lioy and Rappaport (2011). Thinking in terms of the exposome means looking at environmental exposure in a global way. To make this idea clearer, consider the similar movement that happened in genomics, from thinking that single genes could be identified as causes of disease, through realizing the importance of gene-gene interaction, and epigenetics, to studying the whole genome. EXPOsOMICS similarly moves to a global view of disease causation, looking at *total* exposure (internal and external), and also seeking to shed light on gene-environment interaction in disease causation.

Putting the concept of the exposome together with biomarkers research yields the current FP7 project ‘EXPOsOMICS’, which aims to find out a great deal more about disease causation by environmental factors by combining the global approach of the exposome, with the kind of specificity allowed by tracing biomarkers. Their aim is to identify biomarkers for key stages of certain diseases—and particularly of chronic diseases—such as exposure, early clinical changes, and disease development. For example, they might track specific chemicals (e.g., in the air, or in the water of swimming pools) and then track changes that happen at the micro-molecular level internally to the body, by studying biological samples from people. This is ultimately to allow the tracing of the evolution of the disease, and be able to make predictions about diseases, given the biomarkers found in an individual—ultimately building up an idea of how lifetime exposure to certain environmental factors affects individuals. While we are concerned with causes in this paper, it is worth noting that biomarkers are not generally causes of disease. Useful biomarkers will indicate something on the causal pathway, indicating exposure, early clinical changes, or disease development. But some other biomarkers may be mere indicators that the causal pathway is evolving, while not themselves indicating something actually on the pathway. Vineis and Perera (2007) explain:

One of the main challenges with intermediate biomarkers is to understand whether they belong to the causal pathway between exposure and disease, whether they are simply a side effect of exposure or disease, or whether their measurement is confounded by some other exposure. For example, it is likely that certain mutations are genuine intermediate markers in the causal pathway, whereas others are a consequence of the disease, such as genomic instability that arises in cancer cells.

Moreover we will not generally be able to act on biomarkers themselves, but will have to discover where to act. Nelson et al. (2013) discuss the ‘actionability’ of genetic results, but the same arguments apply, *mutatis mutandis*, to biomarkers. Ultimately there are multiple purposes of EXPOsOMICS, tracing the evolution of biomarkers in order to explain and better understand causes of disease, in the hope of ultimately acting on levels of environmental factors to reduce disease.

The vision is to be able to ape GWAS, by conducting EWAS, or ‘exposome-wide association studies’. The idea is that we need to study exposure in all possible forms, particularly, if we can, in the detailed modes of biomarkers research. On the basis of the experience of GWAS, it is suspected that multiple environmental factors will be relevant to many diseases, that there is likely to be interaction between different environmental factors, and of course interaction between environmental factors and genes (see also [Thomas \(2006\)](#)). So the scale of the challenge is well known!

2.2 Discovering and validating biomarkers

EXPOsOMICS is undertaking significant data-driven discovery. Although we know a lot about the body, and about cellular mechanisms, there is no strong general theory of disease causation at the molecular level by environmental factors. Although the mechanisms of action of say, asbestos, thalidomide, or passive smoking have been largely spelled out, the internal exposome still has grey areas. A great deal is still unknown.

In addition to using the omics technologies, which alone yield vast amounts of data, EXPOsOMICS turns to various other sources of data, such as sensor technologies, imaging, electronic diaries, etc., and multiple tools for analysis including high-resolution analytical platforms such as liquid chromatography coupled to mass spectrometry and/or nuclear magnetic resonance spectroscopy. They also use sensors and smartphones to collect measurements about individual physical activity and environmental data (air pollution and level of chemicals in, for example, swimming pool water). Scientists are designing studies where individuals wear palm-size units in back or front packs, connected to their smartphones, allowing GPS and motion sensors to record location of users, and accurate estimates of pollutants in the air.

Scientists are looking for the causal relation flowing from environmental exposure to disease, by measurement of relevant biomarkers to allow the study at different molecular levels of the effect of exposure on environmental agents, and the induced changes in cell metabolism. In the majority of studies, however, we cannot generally intervene on the environment, or manipulate biomarkers of early clinical changes either; in fact one problem EXPOsOMICS tries to address is discriminating between biomarkers that indicate causes of disease and those that indicate effects of the disease. Instead, other studies are called experimental because they recruit participants (e.g., adult, healthy, non-smoking) volunteers (who are informed and give consent) to measure water or air pollution after they have been exposed. In the ‘PISCINA’ study, swimmers in pools are exposed to water disinfection by-products, which are absorbed by the skin and inhaled. The study measures levels of these by-products to assess a) exposure to chlorination by-products in swimmers, b) short-term respiratory health effects and c) genetic damage after swimming. The PISCINA study evaluated biomarkers analysing blood, urine, and exhaled breath condensate ([Kogevinas et al., 2010](#); [Font-Ribera et al., 2010](#); [Richardson et al., 2010](#)).

The outline methodology of EXPOsOMICS is conceptualised as ‘meeting-in-the-middle’, first put forward by [Vineis and Perera \(2007\)](#) and subsequently developed by [Chadeau-Hyam et al. \(2011\)](#). EXPOsOMICS is faced with a complex interacting system, which we are pretty certain consists of multiple interacting causal links. The point of the meeting-in-the-middle methodology is to

probe these unknown causal links, to extract some kind of knowledge of the complex causal structure. Broadly, this involves finding biomarkers of exposure and biomarkers of disease outcome, and *then* finding the ‘intermediate’ biomarkers that link exposure and disease, which are located in the middle of the causal network from exposure to disease.

More specifically, meeting-in-the-middle involves combining the results of prospective and retrospective studies. From prospective studies, scientists extract information about preclinical biomarkers related to particular exposures. From retrospective studies, they extract information that backtracks from clinical disease to preclinical response to exposure. The crucial meeting-in-the-middle bit is in trying to find the overlap, that is, those biomarkers that are good predictors of disease *and* that are associated with exposure. This is when scientists start to link up the evolution of the disease, tracking from cause to effect, for that exposure and that disease. This is the simple story, which will not capture interaction effects. Certainly, EXPOsOMICS researchers use univariate models, comparing factors separately, to measure the marginal effect of each marker. But they also use highly complex multivariate approaches to try to identify the best combination of candidates to predict disease onset, in an attempt to capture interaction effects.

For instance, the work we have mentioned by [Chadeau-Hyam et al. \(2011\)](#) describes a pilot study using data from the European Prospective Investigation into Cancer and Nutrition (EPIC). Researchers compared spectra of plasma samples from 24 cases of colon cancer and 19 cases of breast cancer against 43 controls. Those plasma samples were collected on average 7 years before appearance of cancer. The comparison between cases and controls allowed researchers to identify a putative list of intermediate biomarkers linking exposure and disease.

There are very important design issues for these trials, including when to make the measurements, whether the results of one study are exportable to other contexts—external validity—and identification of false positives and false negatives (see also [Vlaanderen et al. \(2010\)](#)). Nevertheless, the work aims to find causal pathways from exposure to disease, particularly aiming to trace the missing *links* from macro-environmental exposure to micro-molecular changes in the body, and so to the evolution of disease. The thought is that this missing link is given by biochemical pathways. [Thomas \(2006, p490\)](#) says:

Using these “-omics” technologies to inform hypothesis-directed pathway-based approaches to molecular epidemiology and to help direct genome-wide exploratory analyses into more promising directions. [. . .] One might think of the “-omics” data as providing the missing link among exposure, genes, and disease.

The scientists involved are aware of these design problems, which are due to the level of unknowns. They frequently talk of ‘picking up signals’, trying to gather as much relevant data as possible. They make measurements on blood samples before and after the exposure under the hypothesis that there is a short latency to see the effects of pollutants and chemicals. However, they also seek to catch odd temporal effects, so they try to base decisions about measurement on as little biological theory as possible, hoping to strengthen the theory at the same time. That is why in some cases they will opt for 2 measurements, at different times from the exposure event.

The statistics used in biomarkers is very complex and cannot be done with no biological understanding whatsoever. For instance, they build networks to find (explore, validate) the interactions in the exposome (biomarkers of exposure, of early changes, of disease). But statisticians receive row data from the labs, and they have to clean up the data (instrumental errors, normalization, etc.) before analysis. The *statistical* results they obtain need an interpretation from the *biological* perspective. For example, it can make statistical sense to exclude outlying observations, but these may be important from the biological point of view, and mere statistics cannot offer a biological point of view.

2.3 The problems for EXPOsOMICS

In summary, thinking in terms of EXPOsOMICS, using biomarkers, allows epidemiology to take an approach to disease causation by environmental factors that is both more global, in considering multiple factors and the environment both external and internal to the body, and more precise due to using biomarkers. Two processes are tracked: (i) what happens *to* the body when it is exposed to certain chemicals, (ii) what happens *within* the body once the body is exposed.

A great deal is known about such disease causation, but this only serves to highlight what remains unknown. We know we have a complex interacting system of disease causation, involving genetic and environmental factors. We know quite a lot about the system, with many causal links we're perfectly well aware of, and many known mechanisms that we suspect are affected by environmental factors. Previous work, particularly on genomics, has alerted us to the fact that there is a great deal about disease causation still unknown. Previous work also leads us to suspect there will be multiple causes, that have small effects, and strong interaction effects. Previous epidemiological work leads us to suspect that widely divergent kinds of factors will be relevant to disease causation, such as changing social structures that alter people's eating habits, or exposure to pollution, as well as the more obvious exposure to particular substances, or radiation. The EXPOsOMICS project is bringing to bear relatively new omics technologies, alongside data gathered from, for example, smartphones, to make really comprehensive measurements of exposure. This generates a data deluge problem, and interdisciplinary interpretation problems due to the multiple sources of data and different fields involved. The upshot of all this is that projects such as EXPOsOMICS are trying to trace unknown difficult-to-find causal links in the middle of the mess. Yet this is a challenge that must be met if we are to push forward the frontiers of our understanding of disease.

3 Traditional accounts of causal production

3.1 Productive causality: from Hall to EXPOsOMICS

Since Hall (2004), the causality literature has recognised two potential concepts of causality. The first Hall calls 'dependence', but is also often called 'difference-making', and accounts of causality falling in this category include the counterfactual theory, manipulationist and, to some extent, Bayes' nets accounts, which all focus on measuring different types of co-variations between

cause and effect, in abstraction from any particular link between them. The second concept Hall calls ‘production’, and this is the approach we address in this paper. Productive causality is concerned with how causes produce their effects, with how cause and effect are linked.

Whether Hall has genuinely identified two distinct concepts of causation, as he claims, is of course disputed. So is the question of whether a productive account of causality can capture everything that is needed for an account of causality. We tend to the view that you are likely to have to pay attention to both dependence or difference-making, and to production, in order to answer all the questions concerning causality you are likely to want to ask. However, we think that we still need a better understanding of the concept of productive causality itself, even to make progress on the question of how production and dependence may or may not be related. A better understanding of productive causality is our project in this paper.

Existing work has examined the importance of causal production in causal inferential practice in the sciences (Illari, 2011b; Clarke et al., 2012, 2013). There, the core importance of reasoning about causal production is as a way of ruling in and out possible causes, on the basis of what is known of causal linking. This is crucial in helping us to decide which correlations to dismiss, and which to take seriously as possible indicators of unconfounded correlations—as causal relations. This is a very serious problem in any large dataset.

This is a style of reasoning that difference-making accounts of causality do not address. They provide means of ruling out confounding, true, but these are based on further difference-making relations, or means of constraining how we get evidence of difference-making relations, such as randomized or controlled trials. They do not provide any element beyond the difference-making we try to detect in finding correlations, with which to structure the dataset. They do not give an account of how built up background knowledge of causes affects the design and interpretation of current studies.

Note that evidence of production, and of difference-making, are often found in a way that is intertwined. Correlations can provide some prima facie evidence of mechanism, as well as of difference-making. This issue is examined in detail in Illari (2011a); Russo and Williamson (2011); Clarke et al. (2012); Russo (2011), which all examine the integration of evidence of production (specifically of mechanisms) and of difference making in causal inference in medicine and epidemiology. But evidence of production is also sought in understanding means of transmission of influence, such as radiation, or bringing to bear background knowledge of possible causes, in the form of mechanisms already known to link particular causes to their effects.

It is the linking between exposure and disease that EXPOsOMICS scientists are struggling to conceptualise. Consider for example Vineis et al. (2009):

The ultimate goal of using “-omics” technologies to identify environmental causes of disease is to derive an integrated view of the biological processes involved in the continuum from exposure to disease.

They are not merely trying to find out whether C causes E (i.e., whether environmental exposure causes disease, say cancer or allergies), but to discover the *process of evolution* of disease from environmental exposure, to find the connections and map the causal structure—indeed, causal structures. This linking

is measured at the molecular level as explained in the previous section, and results of measurements analysed through complex interaction between statistics, bioinformatics, and biological theory. Given the scale of what is unknown, there is a need coming from the science for a way of conceptualising this linking, conceptualising what EXPOsOMICS scientists are trying to track, in a way that can usefully guide their causal reasoning, meeting the needs identified in section 2.3. A peculiarity of the EXPOsOMICS approach is the *diachronic* assessment of biomarkers. (Vineis and Chadeau-Hyam, 2011, sec.4) explain:

Recently, Sandro Galea has proposed to generalise mathematical approaches initially designed to study infectious disease to chronic diseases epidemiology. In that setting, biomarkers should not be assessed synchronically, as usually done, but diachronically and even ideally their full evolution along time should be considered.

This means identifying the biomarkers of environmental exposure *and* the biomarkers of early clinical changes *and* those of disease. If we can trace the development of these biomarkers, we will then be able to use the right biomarkers for early diagnosis as they will be good predictors of the development of disease.

With this aim in view, we will examine existing accounts of productive causality to see if they can help: processes, complex-system mechanisms, and capacities. Although this last is not strictly an account of causality, it is connected with the aim of illuminating causal linking. For this and further reasons that we will explain, we include it as worth examining. These specific accounts—and accounts of production more generally—have well known problems, which we do not address here. But they also have well known virtues, which make them useful. Space does not allow us to discuss the appropriate place of productive accounts in causal theory, but later we will nonetheless say something about the role of mechanisms in EXPOsOMICS research. Ultimately, we will argue that they are still useful for their known virtues. For now, we will be concerned only with whether they can illuminate linking in EXPOsOMICS. We also lack space to examine variants of the accounts in detail, but this will not be necessary to show the limitations of their application to EXPOsOMICS.

3.2 Processes

The classic approach to productive causality began by looking at physics, in process accounts. The idea of tracing the evolution of the biomarkers echoes process-tracing à la Salmon-Dowe. The ‘combined’ Salmon-Dowe view states that processes are world lines of objects, and *causal* processes are those that transmit conserved quantities (e.g., mass-energy, linear momentum, or charge) after an interaction between two (causal) processes (Salmon, 1984, 1997; Dowe, 1992, 2000). Most recently, this account has been updated by Boniolo et al. (2011). This latest account holds the view that the difference between causal and non-causal processes involves the transmission of *extensive*, not conserved, quantities. This approach is able to account for causation in stationary cases, which the Salmon-Dowe approach could not do.

In many ways, the process tracing accounts set the agenda for philosophical theories of productive causality. One traditional aim they had was to say what causality is, in the world, so they are realist about causality. They also traditionally look to scientific theory and/or practice to find what causality is,

reckoning that science must have something to say about what causality is. We will return to this point in section 4. Since Salmon himself abandoned his mark-transmission account on the grounds that it involved an essential counterfactual element (Salmon, 1994), we do not discuss it in detail here, although we will comment on it further in section 4 because we take mark transmission to be a useful place to start, in order to understand information transmission.

Mark-transmission aside, the Salmon-Dowe process account is clearly tailored to physics, at least in the sense that they use examples from physics only. Indeed, Salmon’s account was designed to discriminate between pseudo and genuinely causal *physical* processes. As such, despite the idea of tracking, and the fact that it was the only such account available for a long time, it doesn’t seem to apply to EXPOsOMICS. The systems described in EXPOsOMICS will presumably obey conservation laws, but we seek more than this. We seek a fine-grained account of the properties in virtue of which environmental factors cause disease, and we seek to give an account that makes sense of tracing those properties through the system from exposure to disease. So the Salmon-Dowe picture is insufficiently illuminating. In EXPOsOMICS, we do not try to track either conserved or extensive quantities.

The broader idea of attempting to track some kind of process is, however, right, and this theory deserves credit for being the only serious attempt to address causal linking available for a long time. It is not by chance that EXPOsOMICS scientists also appeal to concepts that are akin to Salmon’s ‘process’. We aim in this paper to give a better account of the nature of what is tracked, an account that is more informative for EXPOsOMICS.

3.3 Complex systems mechanisms

It is natural to turn next to examine complex systems mechanisms. In the domain of disease causation, scientists frequently use the words ‘system’ and ‘mechanism’ to describe what they’re studying, and background knowledge of the mechanisms and systems of the body is clearly vital to the work of EXPOsOMICS. Broadly, we will adopt the characterization:

A mechanism for a phenomenon consists of entities and activities organized in such a way that they are responsible for the phenomenon. (Illari and Williamson, 2012, p120.)

A great deal is known about the entities or parts of these mechanisms in disease causation—from bodily organs, to DNA—and their activities or interactions—how they work and influence each other. This knowledge is mechanistic knowledge, and it is important to EXPOsOMICS. Witness for instance Vineis and Perera (2007, p1955.):

When combined with the best of the earlier validated biomarkers of dose, effect, and susceptibility, such new markers have the potential to add considerably to knowledge about the mechanistic pathways that relate pathogenic exposures to disease onset and also to serve as informative early markers of disease risk.

Further, mechanisms underlying causally related variables are ways of capturing a kind of linking. However, we will argue that mechanisms are not enough for the

case of EXPOsOMICS. Specifically, the links in mechanisms, while important, are too coarse grained to capture the linking sought.

This is difficult to see, because mechanisms are important in so many ways. They are important in setting up hypotheses for EXPOsOMICS, for suggesting where to look for relevant signs of the action of environmental factors. They can help select what level to carry out omics analyses at—see also later section 4. Witness again from above Vineis and Perera (2007, p1961):

One of the main challenges with intermediate biomarkers is to understand whether they belong to the causal pathway between exposure and disease, whether they are simply a side effect of exposure or disease, or whether their measurement is confounded by some other exposure. For example, it is likely that certain mutations are genuine intermediate markers in the causal pathway, whereas others are a consequence of the disease, such as genomic instability that arises in cancer cells.

Any further things the project finds out about mechanisms will be useful, and it is likely to find out a lot, contributing considerably to our mechanistic knowledge.

Further, what we do know of mechanisms does help us structure data in EXPOsOMICS. Current work in causality (initiated by Russo and Williamson (2007)) has suggested that establishing biomedical causal claims typically requires evidence of mechanisms and of difference-making (correlation in this case). This is broadly because what is known of linking mechanisms helps us decide which correlations might indicate causes, and which might well not.

Unfortunately, known mechanisms are only of limited help for EXPOsOMICS. It is entirely too likely that many of the variables studied are related to many of the other variables studied. This is the challenge of finding causal links in the middle of a highly interactive mess. To put it crudely—EXPOsOMICS scientists already knew an interacting system existed before they started the work! What we know of the mechanisms cannot help us rule out possible causal links in EXPOsOMICS in any useful way.

In fact, it is not always clear how to think of the links in mechanisms. In so far as mechanisms are arrangements of interacting parts, we can understand parts and their activities, but how should we conceptualise linking right across a mechanism? In the mechanisms literature, a distinction is often made between constitutive or vertical mechanisms, which break down a phenomenon into parts, and etiological or horizontal mechanisms, which trace the causal history of a phenomenon or effect. Harold Kincaid explains:

In the philosophy of science literature [...] mechanisms are usually thought of as the component processes realizing some higher level capacities, e.g. the mechanism of memory. I call these ‘vertical mechanisms’. However, many requests for mechanisms are about providing intervening or mediating variables between a putative cause and its effect. Call these ‘horizontal’ mechanisms. (Kincaid, 2011, p73.)

In the practice of mechanism discovery, however, there is no clear distinction between finding constitutive and etiological mechanisms. So when we find parts, how do we find links?

One possible answer is that the overall action of the whole mechanism, entities activities and organization, gives us linking. But in EXPOsOMICS we are definitely looking for quantitative relationships, and building a mechanistic explanation seldom gives you anything *quantitative*. This is not so in all cases (see [Bechtel and Abrahamsen \(2010\)](#)), but even so, there is not enough here.

Another challenge for understanding causal inference using mechanisms is the relation between the single and the general case. We are interested in general-case mechanisms, as is the usual primary case in the literature on mechanisms, and the kinds of mechanisms that are studied extensively in science. Glennan is an exception in the mechanisms literature in treating single-case mechanisms as primary ([Glennan, 2011](#)). However, in the practices of causal inference, it is general-case mechanisms that are used as background knowledge to help structure our inferences going forwards (see [Russo and Williamson \(2011\)](#)). Nevertheless, the relation between the general-case mechanisms we discover and particular causal links is not a simple one. In medicine, for example, finding one mechanism will most often suggest more than one causal link. Frequently, there is a particular causal link when the mechanism is well-functioning, and there may be more than one standard result—an alternative causal link—in different varieties of malfunction.

The core problem here is that mechanisms offer only a coarse-grained attempt to conceptualise some link or other. This is an important role, but biomarkers try to set up a picture of links as straight lines, as opposed to the holism of trying to model the entire system—on this last point see also [Russo and Williamson \(2012\)](#). As biomarkers scientists do, we seek an account of the causal link that can be more precise. We do not suggest that the idea of mechanisms can be replaced in EXPOsOMICS, but that the idea needs to be *supplemented*.

3.4 Powers

We have argued that saying there’s a system where connections take place isn’t sufficient to help with causal inference in EXPOsOMICS, because there’s too much interaction in the system under investigation. What is needed is something more fine-grained. This means that focusing in more narrowly on the powers of parts of that system is worth examining, as these powers are more fine-grained than mechanisms, and so may let us identify something that is more informative.

As with the other metaphysical views we examine, we lack space to consider variants in detail. Here, we focus on the core approach common to powers, capacities and dispositions. The idea is that things have powers to do things. For example, aspirin has the power to cure headaches, when taken in a sensible dose by someone who has a headache. The claim of philosophers such as [Cartwright \(1989\)](#) is that at least some of what science does is study the capacities of things.

This is appealing because the idea that certain chemicals have the power, disposition or capacity to cause certain diseases has intuitive plausibility. The powers of things is also something you do find out about when studying mechanisms. For example, some powers have a mechanistic explanation, such as the power of cells to metabolise lactose in the absence of glucose, which is due to the action of the *lac operon*. And powers are something more specific—more fine-grained—than mechanisms.

Again, finding powers of the parts of the mechanisms of disease causation is

certainly useful. It can be very tricky with a complex interacting system, though, as there is often too much context to attach powers with any real stability to parts of the system. The whole system behaves pretty unreliably. The parts only do their stuff in the context they happen to be in. Stable powers would be great if they could be found, but apart from the very brute things we do know: how we go from DNA to protein, how the circulatory system works in the body, and so on, which of course *structure* a lot of our theorising in EXPOsOMICS, powers are not going to help us with the next phase of discovery.

Seeking powers, if successful, allows the extraction of something useful from complex systems, namely what given entities or systems can or cannot do under specified circumstances. Powers can be useful in explaining causal links. But this is not the only thing we are looking for in EXPOsOMICS, namely the *linking itself*. Powers focus attention on what parts do locally, not on links that stretch across the system. Looking for evidence for powers can help constrain what the activities are, and so what the possible causal links are. But this is not enough. EXPOsOMICS is also looking for very specific, very difficult-to-find, long-range causal connections!

3.5 The limits of available accounts of production

So the problem of EXPOsOMICS is what counts as causal linking in the middle of a complex highly interactive system, where many things could be linked to many others. Their aim is to change the method of assessing exposure to environmental factors, measuring them at the micro, or chemical, level, and detecting biomarkers of exposure at the micro, molecular level. Then, the intention is to trace the evolution of biomarkers up to the development of disease.

The traditional views of production examined above all say something interesting about linking, but all have some limitation in their application to the EXPOsOMICS case. Broadly, the process account is applicable only to physical causality, and is also quite rigid. The mechanistic account is too coarse-grained, and doesn't give a substantive account of what linking itself consists in, particularly in a way that allows for more than one causal link per general case mechanism. The powers account is more fine-grained than the mechanistic account, but does not give an account of linking. We will see in section 5 that these accounts are still useful. But they do not give the whole story for EXPOsOMICS.

What we are looking for instead is an account that is sufficient to capture how fine-grained are the causes sought in EXPOsOMICS. It also has to be very general, so that it can apply in multiple cases, and link profoundly inhomogeneous causal factors, such as chemical agents and social pressures. Finally, if the account is to be informative, to serve the aim of informing reasoning about linking in the science itself, it has to be something that can exist and potentially be found in such a mess of interactions.

It is because it adds these extra requirements for an account of causal linking that EXPOsOMICS is such a nice case to consider in developing a widely applicable account of productive causality. It has noted in the philosophical literature that existing process accounts are not very widely applicable (see for instance Machamer et al. (2000, p7) for biology and Russo (2009, p18) for social science research), but mechanisms and powers have not been examined in the same way. Further, EXPOsOMICS is precisely the kind of case where it is most

tempting to give up on any kind of account of productive causality at all. Yet difference-making accounts do not capture everything that is going on in the reasoning in EXPOsOMICS, as we need reasoning that prioritises some very fragile difficult to find correlations over others. EXPOsOMICS is a special case, but it is far from the only case facing such a large quantity of data, or such an interactive system, while similar projects measuring such divergent factors are sure to follow soon. As a result, an account that offers the prospect of meeting the traditional aims of accounts of productive causality and also meeting the requirements of these kinds of cases should be very interesting both in illuminating reasoning in science, and as a contribution to the philosophical literature on causality.

4 Information

4.1 Picking up signals

Let us start with the hints coming from biomarkers research, their language of “picking up signals”:

While classical statistical models to analyzing -omics data serve the purpose of identifying signals and separating them from noise, little has been done in chronic diseases to model time into the exposure-biomarker-disease continuum. (Vineis and Chadeau-Hyam, 2011, sec.4)

From these two parallel analyses [statistical analyses], we obtained lists of putative markers of (i) the disease outcome, and (ii) exposure. These were compared in a second step in order to identify possible intersecting signals, therefore defining potential intermediate biomarkers. (Chadeau-Hyam et al., 2011, p85)

This is just a hint, and in this section, we aim to spell this out in greater detail in terms of information, and show how by conceptualising causal linking in EXPOsOMICS informationally we can help both the science, and our understanding of productive causality in the causality literature.

In presenting this account of causal production as information transmission, we have three interrelated but distinct aims:

1. Metaphysical: say what causality itself is, starting from EXPOsOMICS science;
2. Conceptual: provide a concept of productive causality that can answer needs that have been recognised in the causality literature;
3. Conceptual: provide a concept of productive causality that can answer the needs of EXPOsOMICS and other similar sciences.

Any of these aims is individually worth achieving. We offer two converging arguments for the view that information is an answer to our question about linking because information is real, and referenced in the scientific practice, in line with the traditional aims of accounts of production. In section 4.2 we argue that causality as information is a natural development of Salmon’s mark

transmission theory. This addresses aim 1 in a manner that is standard in the causality literature. It also addresses aim 2 less directly, although further arguments have already been offered in Illari (2011b) concerning the problem of absences as it arises in the causality literature, which we do not touch on here. We argue that information is naturally the most general thing that can be said about cases of causal production, and explain how this addresses both aims 1 and 2. In brief, information offers both a very general concept of causal linking, and a thin metaphysics of causal production which can be widely applicable, is able to link profoundly inhomogeneous causal factors, and can exist and be found in a mess of interaction.

To be useful, however, the concept we offer will also need to be informative, contentful enough to help solve conceptual problems that arise in EXPOsOMICS. We will show that this is so, by showing how it addresses a particular problem in EXPOsOMICS, the inhomogeneity problem, in section 4.3. This section of course addresses aim 3 directly, but, as we will explain, it also meets the vacuity versus generality problem which is an obstacle to meeting aims 1 and 2. So addressing the inhomogeneity problem completes our arguments.

4.2 Information as the heir of mark transmission

Looking at the problems of applicability that arise in EXPOsOMICS science, alongside these interesting remarks about signalling and the central importance of biomarkers, cannot help but cue the one account of causal production we have not yet considered: Reichenbach-Salmon's mark transmission. This is the theory that Salmon held before the Salmon-Dowe conserved quantities theory, but it is not much discussed, as Salmon ultimately rejected it himself.

The broad idea of mark transmission is that a process is causal if, and only if, were we to mark it, that mark would be transmitted to later stages of the process. So, for example, if we dent a car, the dent is transmitted along with the movement of the car, and we can detect the mark at these later stages. In contrast, if we interrupt the shadow of the car, shining a light, for example, to deform it, that deformation does not travel with the shadow as the car moves, but instead disappears. This distinguishes real causal processes from pseudo-processes such as moving shadows which do not transmit marks. It is also a supremely general account of causal production, as so many different kinds of processes can be marked. The idea makes sense in all of the sciences.

The problem is that many processes are not actually marked and, most damagingly, some physical processes are impossible to mark without fundamentally altering them. Fundamental particles have too few degrees of freedom to allow marks that do not disturb the original process. This means that the mark transmission theory must be given, as we gave it above, as a counterfactual mark transmission theory: *if* a causal process were marked, the mark would be transmitted to later stages in the process. This led Salmon to reject the theory, as he took a counterfactual characterization to be antithetical to the aim of an account of causal production. For EXPOsOMICS, the mark idea is a different kind of problem. Biomarkers are many, and carefully sought. But the causal links sought are often fragile. It is likely that many of the processes could not have a mark *introduced* without deforming the original process.

However, the idea of the mark transmission theory can be recaptured by moving to the idea of *information transmission*. Information is very like a mark

in the relevant sense, as information, too, can and will be transmitted by a causal process. But information does not have to be introduced, as information is *already present* in any causal process whatsoever. The car moving is transmitting information, while the shadow moving is not, whether or not we try to mark either one of them. So the simple formulation: a process is causal if and only if it transmits information of some kind is already as well specified as Salmon's mark transmission account. It shares its major benefit, its applicability, but evades its major flaw. It accords very well with the language of the science in question. It also accords well with Salmon's ideas as he presents them (see also [Weber and Vreese \(2012\)](#)):

It has always been clear that a process is causal if it is capable of transmitting a mark, whether or not it is actually transmitting one. The fact that it has the capacity to transmit a mark is merely a symptom of the fact that it is actually transmitting something else. That other something I described as information, structure, and causal influence ([Salmon, 1994](#), p.303)

A current account of causality in terms of information exists in the literature. It is that of Collier, who holds that causation is the transfer of a particular quantity of information from one state to another ([Collier, 1999, 2011](#)). We shall not further explore his highly technical account of causality in terms of Kolmogorov complexity here, as we will not adopt it. [Ladyman and Ross \(2007\)](#) offer an alternative account of something like informational causality, in terms of Shannon information. From the literature on information in maths and computer science, many formal quantitative measures of information are now available. These give measures that in general can be applied formally to anything whatsoever. In other literature, more qualitative accounts of information are discussed, such as Floridi's account of semantic information ([Floridi, 2011](#)). It is agreed that semantic information is, at the least, supremely difficult to quantify.

The richness of informational concepts is a challenge for an account of causality as information transfer, but it is also its greatest virtue. Collier argues that his aim is to offer an informational account of causality that is general enough to apply to anything whatsoever. His idea is that in more specific contexts, further constraints can be added to the general account—which, as we have said, is already as well-specified as Salmon's mark transmission account—to say something more informative about the special cases. But, necessarily, the most general account can say very little and remain widely applicable. There is a direct trade-off between an informational account being contentful enough to be informative, and it being general enough to be widely applicable.

This problem, however, is not specific to an informational account of production. *It is a challenge for any account of production that aims to be widely applicable.* To make any account applicable right across the sciences will require removing constraints on what counts as a causal link. To make any account informative about particular cases of causal links, or links in any particular domain, requires adding back in such constraints. This is why the richness of informational concepts is a virtue. The very wide applicability of formal information-theoretic concepts gives generality, because absolutely anything can be described informationally using such concepts. However, unlike mark transmission, the formal theories of information, and applications of the idea of information across the sciences, offers theoretical resources that also offer narrower,

but more informative accounts of particular kinds of causal linking. Information itself offers both great generality, and also more informative but restrictive concepts, and resources to offer quantitative measures that may be useful in different cases. No previous account of causal linking has offered *anything* of this kind.

Anscombe (1975) persuaded the causality literature that there is at least a challenge to be faced in the diversity of worldly causes. Is there any way to describe all cases of causality, any thing they all share? We have concepts, both in natural and in scientific language, to describe these causes: binding, bonding, breaking and so on. Is there any feature all of these more specific causal concepts share? Our idea is that all of these are instances, or forms, of information transmission.

In offering an informational account of causality, we are offering, first, a very general concept of productive causality. In the causality literature, authors have often evoked the idea of a ‘secret connection’, following Hume (1777), namely what keeps things together, what links causes and effects, that is the ‘cement of the universe’, to use the language of Mackie (1974). The hope in giving an account of productive causality seems to be that science, or philosophy, will find *the* element that holds the world of our disparate experiences together. But in sciences like EXPOsOMICS, we do not *find* the secret connection, as if it were an extra missing entity in our ontology. In tracing information, in using the conceptual apparatus to re-construct any description of reality in causal-information-theoretic terms, we come to the best understanding we can of what the world is like.

In this way, the concepts we design to understand the world and our conclusions about what we think are in the world are deeply intertwined. This is a neo-Kantian approach, following that of Floridi (2011). In the first place, thinking informationally is a way of designing concepts which, in turn, allow us to grasp important features of the world. What we are doing in offering an informational account of causality in EXPOsOMICS is creating or designing a concept of causal linking that is fit for that domain—and hopefully elsewhere. Coming out of the problem of diversity in the causality literature, we offer a concept that we think can encompass all the rest. The concept of information helps with the conceptualisation of linking of causes and effects. In this way, the extreme generality of the concept of information given to us from mathematical communication theory is an advantage, because information can link inhomogenous factors (see also section 4.3). It is in this sense that information transmission is a thin, a *minimal*, secret connection.

4.3 The inhomogeneity problem

Here we give a positive account of how informational thinking helps in EXPOsOMICS. The idea is that it helps with the *conceptualisation* of links, where the scientists are struggling to conceptualise links.

To say something more substantive about the need for linking, it is worth considering the importance of justifying links that are not immediately plausible. This is a significant concern for scientists. For instance, can mother’s education cause child survival? How? Or, how can obesity cause lack of self-esteem? What are the mechanisms and processes from the social to the biological, and vice-versa? Or what are the mechanisms and processes from the big to the

small, and vice-versa? Causal links between non-homogenous stuff, such as social causes having biological effects, or big things causing small things, are not obviously or non-problematically established in the sciences. The debates about the status of social epidemiology are paradigmatic in this respect: either social factors are ‘genuine’ causes, i.e. they are part of the aetiology of disease, or they are merely ‘extra’ factors to partition the original population of reference. Put otherwise, the question is whether such factors are *causes* or merely *classificatory devices*. Depending on which option one chooses, social epidemiology says, or does not say, something substantial about disease causation. Information, and the reasoning that goes with it, offers a way to make such non-homogenous links not only plausible but also theoretically justified. Analogous reasoning applies to the ‘big’ affecting the ‘small’ and vice-versa.

Consider the links that EXPOsOMICS tries to establish. Environmental exposure is a *macro*-variable that includes coarse-grained data about pollution, and we aim to detect effects at the *micro*-molecular level. Environmental exposure of course consists of pollutants and chemicals we are exposed to. But exposure also needs to be classifiable according to work environment, social class, and so on, to allow us to take the kinds of actions we wish to take, such as policy initiatives. So, environmental factors include biological factors, biochemical factors and socio-economic factors. How can we make the link between environmental factors and disease plausible, in the way that helps EXPOsOMICS scientists support their inferences, *beyond* establishing a ‘coarse-grained’ difference-making relation between the two?

The solution envisaged in EXPOsOMICS is to change the types of measurement used to trace causes. Environmental exposure is assessed by measuring the exact levels of individual chemicals in, say, air or water. Thus finer-grained measurements initially try to restore some kind of *scale* homogeneity: measure the level of a pollutant or of a chemical and then measure levels of changes at genomic, transcriptomic, proteomic, or metabolomic levels. We might think that because ‘scale’ homogeneity is restored through making all measurements biological measurements, the problem is solved. Not quite.

One reason is that measurements now taking place at the same level allow EXPOsOMICS scientists merely to establish another difference-making relation, but at a much lower level. For instance, they could establish a robust difference-making relation between the level of a certain chemical in swimming pool water and the biomarker of early clinical changes of a targeted disease. But this doesn’t establish a *link* yet. On the one hand, they need to find the *right* ‘intermediate’ biomarkers, the ones that are linked to exposure and to disease. On the other hand, they need to place this *reconstructed* link into a plausible network of relations. Recall, the internal exposome is anything that happens inside the body as a response to external exposure. Whether scientists hit upon the right intermediate biomarker will be theoretically justified to the extent that the complex (internal) biochemical mechanisms also include that biomarker. This means that linking cannot be seen with the naked eye, and not even using experimental set-ups. Instead, linking is reconstructed by putting together the many pieces of the evidential puzzle. This requires a lot of empirical evidence and a lot of *interpretation* of the evidence using the right concept. The thought is that information is precisely one concept needed to do that. It is worth noting that this problem is not specific to EXPOsOMICS. It is shared by experimental and observational methods alike. In fact, any scientific conclusion is the result

of a reconstruction and interpretation of evidence.

A second, more important, reason that the problem is not solved is that although homogeneity in the scale of measurement is restored by using biological measurements, this actually makes the results harder to interpret, because the interpretation will still have to be at the macro level, i.e. the level of environmental exposure causing disease. On the one hand, there is a lot of uncertainty and discussion within EXPOsOMICS projects about the level at which biomarker analyses have to take place. EXPOsOMICS scientists talk about targeted omics, when they know at what level—e.g. metabolite—to look for an effect, i.e., they know already from previous research what biomarkers to look for. They talk about untargeted omics when they don't know when or where things will show up. Biological theory often *suggests* what level to run omics analyses at, or when to collect samples, but biological theory is also trying to understand mechanisms better and thus support EXPOsOMICS. On the other hand, the homogeneity problem fires back: you measure everything at the micro-level, level of pollutant, and level of metabolite. But ultimately what you want to know is how and to what extent environmental pollutants cause cancer, and cancer is a much bigger thing than a change in the level of something in a cell. So everything will need to be re-interpreted at the macro-level.

More generally, any causal claim derived from biomarkers research will be an *interpretation* of very many pieces of the 'evidential' puzzle. It will be a *reconstruction* of information coming from biomarkers analysis, plus statistics, plus biological theory. It will be an a posteriori reconstruction of data- or technology-driven research. In this context, informational thinking helps with conceptualising production (the linking) as the evolution of biomarkers, from exposure to early clinical changes, to disease. Informational thinking provides the conceptual framework for the evolution of the process of disease, to be reconstructed by putting together various pieces of the evidential puzzle and by tracing the transmission of relevant marks.

In less complex—or better known—mechanisms or systems, scientists build up a view of the typical kinds of activities and entities in their fields, and the typical powers of the entities. For example, mRNA binds to the ribosome, to allow tRNA to bring amino acids and arrange them according to the code on the RNA strand to form a polypeptide chain. Scientists come to understand their classic entities, the DNA and the tRNA in this example, and what they do, their characteristic activities, the binding, bringing and arranging in this example. This means scientists working in a particular field begin to get a feeling for what happens, and what can happen—they get a feel for causal linking. In a system of the kind faced by EXPOsOMICS, where the vast majority of interactions are so completely unknown, scientists cannot rely on a pre-existing feel for possible linking. Further, when a team consists of specialists in biochemistry, right through to population-level epidemiology, the characteristic entities and activities of their fields do not easily link up, in the way that scientists working on protein synthesis get a feel for how bringing and arranging amino acids ultimately creates proteins. But how do we intuitively link up chemical binding, and smoke levels in the atmosphere? Each field knows only its own activities and entities, has little feel for those of other fields, and since this is frontier research, nobody knows yet how they will integrate. Information lets us conceptualise links in the discovery process, before we have detailed mechanistic knowledge.

Let us wrap up: traditional epidemiology established difference-making rela-

tions between macro-variables describing the environment and macro variables describing diseases by establishing robust correlations (difference-making relations) between them. But we want to know more: what exactly is in environmental factors that causes e.g. cancer? So EXPOsOMICS scientists go much smaller: they measure pollution at the chemical level, measure effects of pollution at the molecular level, using biomarkers. Does this establish a *link* between micro-environmental factors and micro-disease? We argued that it doesn't because this would just establish another difference-making relation, but at much lower level. So the problem is not one of scale of measurement: the problem isn't solved just by making measurements at the micro-molecular level. To establish the link we need to find the *continuum* from exposure to disease, via key stages of the evolution of biomarkers. What we have to establish is continuous linking, not just difference-making at a different level of measurement.

So if the homogeneity is not what EXPOsOMICS scientists track, what is it? *Continuity* in linking is sought, which is reconstructed from the interpretation of tons of data through very different analyses and using many fields of expertise. Biomarkers research tries to find the evolution of biomarkers, thus suggesting a continuity in the process from exposure to disease, and biomarkers help detect some key stages. The idea is to avoid having a rigid distinction between kinds of relata at different scales—a rigid distinction between macro causes and effects, and micro causes and effects. The 'flow' of the process of evolution of disease is not captured by an entity-activity ontology, as used in mechanisms (see discussion of mechanisms above).

Ultimately, we want to understand the whole system, all the bits, how they interact, quantitatively, build reliable models of the dynamic evolution of whole systems under many many different exposure conditions, and the concept giving the dynamic evolution is information. The flow is in the link, and the link, we suggest, is given by information. More precisely, it is given by the reconstruction of the information flow through the different types of analyses. The challenge is how to represent and conceptualise these diverse kinds of things in a way that is a) general enough to see similarities and differences and b) not vacuous, contentless. So it needs to constrain our thinking in some significant ways—constrain it to enable it (see also the next section on this point).

Now it is true that in the idea of exposome-wide association studies (EWAS), information is still not well-defined. In GWAS, there is at least some possibility of a quantitative definition of information, as genes are much more clearly defined than environmental factors, and substantive informational concepts make sense when applied to genes. However, as we have said, the diversity and richness of informational concepts, which appears to be a weakness of an informational approach, is in fact a virtue. There are, for example, qualitative notions of information that could be applied here. As we have argued, though, the very general, thin idea of information that we have offered is already helpful in conceptualising linking in EXPOsOMICS.

5 Back to traditional accounts of causal production: mechanisms as information channels

In pointing to the limits of available accounts of causal production, we did not mean that they have to be altogether abandoned in favour of a purely informational approach. We think that an account of causal production as information transmission can usefully be *added* to our stock of accounts of causal production. We are not advocating that informational thinking can replace the useful work done by the other accounts. In particular, we think information synthesizes very nicely with accounts of mechanisms, as a way of understanding EXPOsOMICS research. We also think conserved quantities and capacities, powers and dispositions are useful in various ways, although we do not address them further in this paper.

We mentioned earlier that, in the causality literature, there is work showing that mechanisms are important in causal inference, not least in that evidence of mechanisms integrates with evidence of difference-making in establishing causal relationships, particularly in medicine (for a summary of these views, see [Clarke et al. \(2012, 2013\)](#)). We agree with that view. We will here explicitly address the connection between this current work on EXPOsOMICS and this previous work.

Broadly, we take it that the two views are completely compatible if we understand *mechanisms as information channels*. The thought is to understand mechanisms as the channels through which information flows. The entities, activities and organization of the mechanism set the channel conditions that allow the flow of information. What we know about these entities, activities and their organization helps us in finding the more fine-grained causal links we seek. This is how mechanisms can yield coarse-grained causal linking, which can usefully be supplemented by fine-grained linking. The structure of the mechanism, though, is often easier to detect, easier to influence, and so on, which means that the coarse-grained linking of a mechanism cannot be simply replaced by a finer-grained account in terms of information flow.

The following metaphor can be used to illustrate this. Production—the linking—is a branch floating down a river. Production happens in some context, which can be thought of as the river. With the scientific study of processes, mechanisms, or capacities or powers we find and define the banks, and perhaps some of the currents of the river. All this helps track the branch in the river, including intercepting the branch again after it bobs under the water and can't be seen by us for a time. Along the river there are wood factories where we can process branches to produce inferences, explanations, predictions, etc. The banks of the river are the mechanism, and they constrain and structure the flow of the water, which is the information. The flow of the water can be changed by directly removing or adding water, or dyeing it orange, or floating a boat on it. But it can also be changed by altering the banks of the river. Both are useful.

This accords with the extremely important understanding of life as essentially something that can resist thermodynamic entropy. Life stays organized, by taking in energy, when inanimate things, which cannot take in energy by themselves, inevitably degenerate. In succeeding in this, living things have to resist environmental variation, such as variation in temperature. Many living things, including cells and multicellular organisms, achieve this precisely by maintain-

ing an internal environment that suits them, and resisting external changes. Interestingly, this is the same importance of the internal and external environments that EXPOsOMICS recognises as the total exposome. Many many of the mechanisms of life exist to allow organisms to do this, i.e. specifically to support some kinds of informational transfer, such as food is allowed in, and proteins urgently needed are made, and resist others, so that extreme temperatures remain outside, with internal temperatures remaining fairly constant, and invading viruses are not allowed to fulfil their purpose, but are rapidly killed.

This idea also helps address two problems in the mechanisms literature. The first is the relationship between single-case and general case mechanisms. We have said that we will concentrate on general-case mechanisms. Certainly, as science is practised, particularly as causal inference in the sciences proceeds, what matters are general mechanisms. (The process of mechanism discovery is discussed extensively in [Darden \(2006\)](#).) Those are the mechanisms that we find, that then feed into our background causal knowledge, and we use to structure our thinking about possible causal links. But we can see now that particular causal links are particular transmissions of information across a mechanism. This helps address the problem [Kincaid \(2011\)](#) raises of the difference between horizontal and vertical mechanisms alluded to earlier in the paper. Vertical mechanisms are the understanding of the banks of the river, the mechanism's parts and so on, across which causal influence can be transmitted.

The second problem, not much remarked on, is that mechanisms tend to maintain more than one causal link. Standardly in the medical literature, for example, there will be the characteristic upshot of a well-functioning mechanism, and of the same mechanism malfunctioning, perhaps in two or more definite varieties of malfunction. This can be characterised as correct information transmission and information transmission misfire in ways that traditional accounts of mechanism struggle with. This will enable a significantly more nuanced account of the use of mechanisms in causal inference.

6 Discussion and conclusion

Biomarkers research is an emergent field of research at the crossroads of epidemiology, molecular epidemiology, informatics, biology, and statistics. There are many innovative aspects to this research. Some concern the methodology, such as the 'meeting-in-the-middle' methodology or the use of omics technologies. Some others are conceptual, such as the design of the concept of 'exposome'. EXPOsOMICS research does not only need to design 'new' concepts, such as the exposome, but also to redesign existing concepts, most notably 'process'.

EXPOsOMICS scientists aim to trace a particular process: the evolution, development, or progress, of biomarkers, from exposure to early clinical changes to the development of disease. But tracing such a process is very different from what philosophers of causality have previously envisioned as processes, designed using examples from physics and space-time philosophy. To trace disease processes using biomarkers, scientists have to carry out numerous analyses of data, from different disciplinary perspectives. Further, the evolution of biomarkers is the upshot of many factors operating at different levels. And the process to be discovered will provide the 'missing link' between environmental factors and diseases.

So how can we make sense of this process tracing in EXPOsOMICS? In the paper we examined traditional accounts of productive causality (processes, mechanisms, capacities) to see whether they can provide a concept of linking that meets the challenges we highlighted. We argued that these approaches have serious limitations in their application to EXPOsOMICS, either because they provide a concept of linking that is too specific, due to being tied to one domain; or because they provide a concept of linking that is too coarse-grained to help with thinking about EXPOsOMICS.

We then investigated the prospects of an *informational* approach, offering two converging arguments. First, we argued that thinking of causal production as information transmission should be seen as an adaptation of Salmon's mark transmission account, which inherits its wide applicability but avoids its main weakness as it does not need to be characterized counterfactually. We explained how we see this offering a very generally applicable concept allows us also to express what we have most reason to believe is in the world. We also argued that the richness of informational concepts is a virtue of the approach, rather than a vice.

Our second, related, argument was that the thin idea of thinking of causal production informationally was already enough to help with some conceptual needs in EXPOsOMICS. This need is conceptualizing the linking of non-homogeneous factors. How can 'macro' factors such as environmental factors have effects at the micro, molecular level? We explained that this question cannot be answered merely by changing the scale of measurement. An answer requires conceptualising *continuous linking* between different types of causes and effects, at different scales. Continuous linking is established by EXPOsOMICS scientists via an interpretation of the complex evidential puzzle. Consequently, information transmission will not be something extra that we see 'out there', another entity in our ontology. Continuous linking is this minimal metaphysics that we construct from the evidential puzzle.

Much has to be done to build a full-blown informational approach to productive causality. But we hope that we have placed the first corner stones of such an account. We also hope to raise interest in information, and, in using scientific research in progress, we show how philosophy and science can work together on timely issues of common concern.

Acknowledgements. We are deeply indebted to Paolo Vineis and collaborators on the EXPOsOMICS project at Imperial for discussions on these issues. We gratefully acknowledge the financial support of the FWO-Flanders (F. Russo has been a 'Pegasus Marie Curie Fellow' during the academic year 2012-2013) and the AHRC (P. Illari has been working on the project 'Understanding information quality standards and their challenges' at the University of Hertfordshire, 2011-2013). We also thank audiences at the Fifth Workshop on the Philosophy of Information, University of Hertfordshire 2013; the "Causality and Mechanisms" workshop of the Egenis group, University of Exeter 2013; the Causality and Experimentation in the Sciences Conference, Sorbonne, Paris, 2013; and the European Philosophy of Science Association, Helsinki, 2013, especially Erik Weber, and one anonymous referee, for useful comments. Remaining errors are, of course, our own.

References

- Anscombe, G. (1975). Causality and determination. In *E. Sosa, (ed.) Causation and Conditionals*. OUP.
- Bechtel, W. and Abrahamsen, A. (2010). Dynamic mechanistic explanation: Computational modeling of circadian rhythms as an exemplar for cognitive science. *Studies in History and Philosophy of Science*, 1:321–33.
- Boniolo, G., Faraldo, R., and Saggion, A. (2011). Explicating the notion of ‘causation’: the role of extensive quantities. In Illari, P., Russo, F., and Williamson, J., editors, *Causality in the sciences*, chapter 24. Oxford University Press.
- Cartwright, N. (1989). *Nature’s Capacities and their Measurement*. Clarendon Press.
- Chadeau-Hyam, M., Athersuch, T. J., Keun, H. C., Iorio, M. D., Ebbels, T. M., Jenab, M., Sacerdote, C., Bruce, S. J., Holmes, E., and Vineis, P. (2011). Meeting-in-the-middle using metabolic profiling—a strategy for the identification of intermediate biomarkers in cohort studies. *Biomarkers*, 16(1):83–88.
- Clarke, B., Gillies, D., Illari, P., Russo, F., and Williamson, J. (2012). The evidence that evidence-based medicine omits. *Preventive Medicine*.
- Clarke, B., Gillies, D., Illari, P., Russo, F., and Williamson, J. (2013). Mechanisms and the evidence hierarchy. *Topoi*, forthcoming.
- Collier, J. (1999). Causation is the transfer of information. In Sankey, H., editor, *Causation, natural laws, and explanation*, pages 215–263. Kluwer, Dordrecht.
- Collier, J. (2011). Information, causation and computation. In Crnkovic, G. D. and Burgin, M., editors, *Information and Computation: Essays on Scientific and Philosophical Understanding of Foundations of Information and Computation*. World Scientific, Singapore.
- Darden, L. (2006). *Reasoning in Biological Discoveries*. Cambridge University Press, Cambridge.
- Dowe, P. (1992). Wesley Salmon’s process theory of causality and the conserved quantity theory. *Philosophy of Science*, 59(2):195–216.
- Dowe, P. (2000). Causality and explanation: review of Salmon. *British Journal for the Philosophy of Science*, 51:165–174.
- Floridi, L. (2011). *The philosophy of information*. Oxford University Press.
- Font-Ribera, L., Kogevinas, M., Zock, J.-P., Gómez, F., Barreiro, E., Nieuwenhuijsen, M., Fernandez, P., Lourencetti, C., Prez-Olabarra, M., Bustamante, M., Marcos, R., Grimalt, J., and Villanueva, C. (2010). Short-term changes in respiratory biomarkers after swimming in a chlorinated pool. *Environmental Health Perspectives*, 118:1538–1544.
- Glennan, S. (2011). Singular and general causal relations: A mechanist perspective. In Illari, P. M., Russo, F., and Williamson, J., editors, *Causality in the Sciences*, pages 789–817. OUP, Oxford.
- Hall, N. (2004). Two concepts of causation. In Paul, L., Hall, E., and Collins, J., editors, *Causation and Counterfactuals*, pages 225–76. MIT Press.
- Hume, D. (1975 (1777)). *An Enquiry Concerning Human Understanding*. Clarendon Press, Oxford, posthumous edition edition. Part II, Paragraph 60.
- Illari, P. M. (2011a). Mechanistic evidence: Disambiguating the Russo-Williamson thesis. *International Studies in the Philosophy of Science*, 25:139–157.
- Illari, P. M. (2011b). Why theories of causality need production: an information-transmission account. *Philosophy and Technology*, page 20.
- Illari, P. M. and Williamson, J. (2012). What is a mechanism?: Thinking about mechanisms across the sciences. *European Journal of the Philosophy of Science*, 2:119–135.
- Kincaid, H. (2011). Causal modeling, mechanism and probability in epidemiology. In Illari, P., Russo, F., and Williamson, J., editors, *Causality in the Sciences*. Oxford University Press.

- Kogevinas, M., Villanueva, C. M., Font-Ribera, L., Liviach, D., Bustamante, M., Espinoza, F., Nieuwenhuijsen, M., Espinosa, A., Fernandez, P., DeMarini, D., Grimalt, J., Grummt, T., and Marcos, R. (2010). Genotoxic effects in swimmers exposed to disinfection by-products in indoor swimming pools. *Environmental Health Perspectives*, 118(11):1531–1537.
- Ladyman, J. and Ross, D. (2007). *Every Thing Must Go: Metaphysics Naturalized*. OUP. With David Spurrett and John Collier.
- Lioy, P. J. and Rappaport, S. (2011). Exposure science and the exposome: An opportunity for coherence in the environmental health sciences. *Environmental Health Perspectives*, 119(11):a466–a467.
- Machamer, P., Darden, L., and Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science*, 67:1–25.
- Mackie, J. L. (1974). *The cement of the universe. A study on causation*. Oxford University Press.
- Nelson, N. C., Keating, P., and Cambrosio, A. (2013). On being ‘actionable’: Clinical sequencing and the emerging contours of a regime of genomic medicine in oncology. *New Genetics & Society*, page in press.
- Rappaport, S. (2011). Implications of the exposome for exposure science. *Journal of Exposure Science and Environmental Epidemiology*, 21:5–9.
- Richardson, S., DeMarini, D., Kogevinas, M., Fernandez, P., Marco, E., Lourencetti, C., Ballest, C., Heederik, D., Meliefste, K., McKague, A., Marcos, R., Font-Ribera, L., Grimalt, J., and Villanueva, C. (2010). What’s in the pool? A comprehensive identification of disinfection by-products and assessment of mutagenicity of chlorinated and brominated swimming pool water. *Environmental Health Perspectives*, 118:1523–1530.
- Russo, F. (2009). *Causality and causal modelling in the social sciences. Measuring variations*. Methodos Series. Springer, New York.
- Russo, F. (2011). Causal webs in epidemiology. *Paradigmi*, XXXIX(1):67–98.
- Russo, F. and Williamson, J. (2007). Interpreting causality in the health sciences. *International Studies in the Philosophy of Science*, 21(2):157–170.
- Russo, F. and Williamson, J. (2011). Generic vs. single-case causal knowledge. the case of autopsy. *European Journal for Philosophy of Science*, 1(1):47–69.
- Russo, F. and Williamson, J. (2012). Envirogenomarkers. the interplay between difference-making and mechanisms. *Medicine Studies*.
- Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton University Press, Princeton NJ.
- Salmon, W. C. (1994). Causality without counterfactuals. *Philosophy of Science*, 61:297–312.
- Salmon, W. C. (1997). Causality and explanation: a reply to two critiques. *Philosophy of Science*, 64(3):461–477.
- Thomas, D. C. (2006). High-volume “-omics” technologies and the future of molecular epidemiology. *Epidemiology*, 17(5):490–491.
- Vineis, P. and Chadeau-Hyam, M. (2011). Integrating biomarkers into molecular epidemiological studies. *Current Opinion in Oncology*, 23(1):100–105.
- Vineis, P., Khan, A. E., Vlaanderen, J., and Vermeulen, R. (2009). The impact of new research technologies on our understanding of environmental causes of disease: the concept of clinical vulnerability. *Environmental Health*, 8(54).
- Vineis, P. and Perera, F. (2007). Molecular epidemiology and biomarkers in etiologic cancer research: The new in light of the old. *Cancer Epidemiology, Biomarkers & Prevention*, 16(10):1954–1965.
- Vlaanderen, J., Moore, L. E., Smith, M. T., Lan, Q., Zhang, L., Skibola, C., Rothman, N., and Vermeulen, R. (2010). Application of OMICS technologies in occupational and environmental health research; current status and projections. *Occupational and Environmental Medicine*, 67:136–143.

- Weber, E. and Vreese, L. D. (2012). Causation in perspective. are all causal claims equally warranted? *Philosophica*, 84:123–148.
- Wild, C. P. (2005). Complementing the genome with an “exposome”: The outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer epidemiology, biomarkers & prevention*, 14:1847–1850.
- Wild, C. P. (2009). Environmental exposure measurement in cancer epidemiology. *Mutagenesis*, 24(2):117–125.
- Wild, C. P. (2011). Future research perspectives on environment and health: the requirement for a more expansive concept of translational cancer research. *Environmental Health*, 10(Suppl 1).