7

A New Economics of Attention?

Introduction

In *Travelling Representations*, we were concerned with attempts to use sociological findings as premises for making epistemological claims about scientific practices. Here, we turn to the scientists themselves and their reactions to the descriptions and diagnoses philosophers and sociologists are promulgating. What we see in their discussions is not a set of disciplines flummoxed by interest driven puzzles about philosophical questions, but practical members of project teams trying to respond to the requirements of a new economics of attention. Attention is always a scarce resource in any walk of life and its allocation has to be budgeted in terms of the opportunity costs of attending to somethings rather than others. A recent colloquium, *The Natures of Data* [Fischer et al. 2020], brings out how the economics of attention may shifting in one corner of the scientific world and how those shifts are being experienced and rationalised. The volume is a reconstructed conversation between a group of biologists, an historian of science and a video-artist interested in scientific practice. The topics cover various aspects of the consequences of the current 'digital revolution' taking place in the biological sciences. Important new research procedures and processes are being developed while at the same time entirely new forms of data and new ways of handling old forms of data are emerging. These new practices are bringing about the shifts mentioned above.

The materials the scientists discuss are fascinating, providing new and different descriptions of actual scientific work. What is less fascinating (dispiriting, even) is the rationalisation which seems to be taking hold. This is couched in a Latourian post-modern idiolect in which the central pillars of what it is to do biological science are held to be under threat, prime among which are said to be the very facticity of flora, fauna, species and other key ontological categories.

Set aside the apocalyptic clamour of the applied sociology and focus on the actual work the scientists discuss and things look somewhat different. True, the scientists are having to 'retool' their research teams, accommodate new specialisms and form new partnerships. As any established researcher knows, these are not comfortable things to have to do. They take time, effort and resources (especially attention, hence our emphasising it), which one might better spend getting on with the science one wants and knows how to do. However, the discombobulation of change is not the same as the collapse of the scientific world as the scientist knows it. Only a hyperbolic sociology would see it as such. When looked at from other more or less conventional sociological viewpoints, what is going on seems entirely predictable and reassuringly normal.

Section 1. An Aside.

The Natures of Data is an output; a project product where a group of scientists (or researchers, if you want to quibble about the scientific credentials of the video artist) have co-constructed a discussion document (literally) about their science. One possible question which could be asked is about its character as a research object. What are its features as a course of biological (or meta-biological, if you will) reasoning in action? Asking this question brings out the very obvious fact that, as data, it is very different to the standard ethnographic material used in studies of science. It is an account the scientists themselves have produced about what they do and what they think it means or might mean. Drawing a parallel with another famous study, Garfinkel's Pulsar Paper [Garfinkel 1981] was about treating a transcript of the work a group of scientists as the collective making visible of their working out of the astronomical meaning of the readings they were getting. *The Natures of Data* could be construed as the reconstructed logic of a similar exercise, the authors making visible their working out (or, at least, attempts to work out) the biological significance of some of the things they are now doing. Adopting this frame of reference forces attention onto the structures this 'research product' might display and how they might be organised. What are the requirements of/the constraints on recipient design placed on the descriptions and on the readings the audience should make as well as who makes up their audience. How is this presentation of their talking to each other shaped as a 'presentation' and for which types of anonymous others? How are these social types discernible from what they say and how they say it? How does the text provide for the self-selection of appropriate audiences and recipients? Most importantly of all, perhaps, how are these to be made sociologically visible and available? What repertoires of formulation and their notations are available to analysts of these discussions and how are the contributions the scientists make to be rendered using them?

Treating the research products as constructed forms of 'instructed reading' might open up all sorts of interesting lines of investigation. Some of them might even bear on the vexed questions of multi- and cross-disciplinary research. Hints of how this might go do appear here and there in the discussion below but undertaking such an analysis is not what we are about.

Section 2. What's Different about the Digital?

The story of Henny Penny amuses because of the universality of the trait it exemplifies. We all know people for whom any large-scale disjunction/disruption/reconfiguration constitutes wholesale revolution portending termination of 'the ways things are'. And, we have to admit, most of us have some of that inclination too. In the past, we've talked about it as the fallacy of the immediate consequential on the shock of the new. What Henny Penny (and the rest of us in our het-up moments) fail to recognise is just how much of what we do and how we live is *not* affected by whatever it is that is disturbing us right now. Covid-19 brought disruption, anxiety and difficulty. But it has not brought the end of the world. We have adjusted. We are adjusting. And while some things might be different for some time, most other things will continue pretty much as they were. Houses will be decorated; children will be brought up; tv will be watched; arguments will be had over who left the light on overnight or the fridge door open.

Precisely the same is true in the development of science. The revolutions in instrumental technologies of the 17th, 18th, 19th and 20th centuries did change the way the sciences were practised. Digital technologies are doing the same today. But what changed the character of the sciences *as sciences* were the ideas not the technologies. And the Kuhnian in all of us should force a recognition that post Galileo and Newton, we have really had only two revolutionary ideas, natural selection and quantum theory. Einstein's special and general relativity weren't really revolutionary since they were attempts to work within the classical conception and the limitations of so doing generated the break-through attributed to him. It is a well-worn story that Einstein could never really reconcile himself to the consequences of his own work.

But even those two 'revolutions' didn't change *everything*. Since the sciences are path dependent, a great deal of how you do practical Biology and practical Physics carried on much as before. Nonetheless, the path had altered and the consequences of these alterations gradually worked their way out. This is where the issues of what is significant and for whom arise. Philosophical and sociological significances have to do with philosophical and sociological questions and puzzles. And these questions do not rest on the particularities of any local forms of scientific practice or theoretical knowledge. Conceptions of ontological differentiation are not overturned because one can describe the behaviour of light in equations written for waves of radiation instead of pulses of particles. Conceptions of causal ordering are not jettisoned because our parameterisation of time yields backward running clocks. And, perhaps most important of all, the metaphysical character of 'an object' is not dissolved simply because the objects we are now discussing can be seen only via complex, highly sophisticated and not always well understood machines and their 'invisible' processes. We can talk about quarks, genes and even "digital fish" (if we must) without becoming unhinged.

The upshot of all of this is that what are posed as all-embracing philosophical issues usually turn out to be much more low-key ones of accommodation and adjustment. How do the things we want to say about the new science fit in? What needs to be flexed, re-thought, re-phrased as well as jettisoned? Revolutionary changes in our philosophical analyses don't come from the outside, from the topics/examples being discussed—particularly those of the sciences, but from within; from philosophers who radically change the questions we are asking, often by pointing out those questions are actually nonsensical.

The implications for sociological analysis are much the same. The sociological conception of human action operational in any mode of analysis is not totally transformed by the contents of the study. Merton would have said the same things about Genomics as he did about 19th century Mechanics (and so will SSST researchers). Some of the participants in *Natures of Data* occasionally lapse into Latourian idiolect, even though ANT itself is little more than the use of network imagery to extend Critical Theory. Insisting the embedded institutionalisation of ways of thought and practice plays out as the preservation of the interests of the powerful is just a tactic for making something quite ordinary and well understood sound mysterious and dangerous. So, more relevantly for the case in hand, is the notion of 'de-materialisation'. Prodded by the 'problematising' strategy of the non-biologists in the group, the scientists themselves seem to have been persuaded (at least for this occasion) to entertain such hyper-inflated notions as the 'de-materialisation' of phenomena, infrastructural 'actors', 'opaque' software, and miscegenated digital monsters.

These terms are a mechanism for dramatising the significance of the issues being discussed. Here are two quotations from the discussion of infrastructure which demonstrate what we mean.[1]

> First:
>
> **Hannes Rickii**: .... The question is: Do these elements contribute at all to shaping research work, and how does this emerge in the context of experimental work? Are infrastructures and nature considered factors of uncertainty, or do these peculiarities get lost in the normality of research work? [p.97]
>
> And then:
>
> **Hans-Jorg Rheinberger** On the whole, however, there is only one kind of materiality involved here, namely the materiality of technology. But there is also the materiality of the object of investigation. It disappears here, it is not present at all, no fish, no nothing at all. No genome. Nothing.
>
> **Hannes Rickii** The fish is there.
>
> **Hans-Jorg Rheinberger** Where is it? Here, in the aquarium?
>
> **Hannes Rickii** Precisely. And that is my motive. I realize that with development toward digital science, much more infrastructure comes between the observer and the object of research. My thesis is that the economy of attention is distributed differently. The materiality of the research enterprise is no longer related exclusively to the investigated organism. Thus, the focus is shifting to what materially makes up the thread between the animal subject and the observer. [pp 101-2]

What is being discussed here is the interpolation of various technologies such as sensors, communications devices (together with image processing, data compression, storage devices and algorithms to run them) between the object being investigated (fish in the Arctic Ocean) and the researchers in their labs. What they are bumping up against are issues of auditing the research activity. Where are the dependencies? Where is what effort expended? What do they have to spend time worrying about? Each aspect of the discussion (data, infrastructure, software, in silico etc) comes at that problem from a slightly different direction and has different consequences for "the economics of attention".

---

[1] All unattributed quotations are to [Fischer et al. 2020]

Section 3. Practical Genomics

The best and certainly the longest example of the above tendency is the section labelled *In Silico*. It is worth spending a little time sketching an alternative version of what is described in this part of the book just to show it is possible to give a sociological account of what is being described without losing one's balance. The case is a familiar one to introductory Biology students (or so we are told). In many if not most species, embryos develop through three stages. Species diversification is evident in stages 1 and 3. In stage 2, however, there is a high prevalence of a very small number of forms. While there are two competing 'hypotheses' offered for this pattern, no universally accepted explanation has been provided. The pattern is referred to as 'the hourglass problem'. One potential clue, however, has been spotted in the high correlation of the 'gene expression profiles' in stage 2. One of our scientists, Hans Hofmann, has been trying to solve the hourglass problem using Genomics.[2]

A major industry in Genomics is the identification and tracking species' genes as individuals develop. This is called 'transcriptomics'. The researchers engaged in this work develop large databases of 'gene transcripts' which list the order of the key bases (labelled A, C, G, T) which make up genes for the species they have studied. What Hoffman wants to do is identify the clusters of common genes for the species being investigated ('orthologous genes') in the critical stage 2. He looked at 2,500 genes to see if he could find such clusters. Basically, he threw statistics and simulation at the problem; a strategy which he could not have used before high-speed desk-top computing and large-scale storage were available. (The point about desk-top computing will emerge later).

The first step is to conduct a Principal Components Analysis. What Hoffman has (or will have when he shapes up the data – see below) is a database of 20 samples of 2500 genes x 15 sub-stages of development containing the A, C, G, and T values for each stage for each gene.[3] The matrix of correlations of the order the bases take measures how distinct or 'distant' each pair of species is from each other. This is a high dimension matrix and impossible to scan visually or parse by hand. The aim is to reduce the dimensions to a low (that is, handleable) number. What PCA does is rotate the matrix by zeroing on the centroid of the distribution (and thus re-calibrating the co-ordinates of the genes). Each rotation reduces the dimensionality of the matrix by 1. Clearly,

---

[2] The scientific detail is presented and its formulation as a testable hypothesis are presented on pp 118-9.

[3] Hofmann doesn't define the matrix but talks of "hundreds of millions of data points". This is our guess.

this process can be iterated until only one dimension is left, but that would squash all the data onto a single line and throw away too much information. Instead, what researchers try to find is smallest number of rotations 'explaining' the highest proportion of the variance in the matrix. The assumption is all 2500 genes contribute to the variance and if you cluster them by clumping them in this way you will separate out grouped contributions to that variance. Following normal practice, Hofmann boils the data down to three components which, using the components as coordinates, allows convenient Cartesian visualisation of the genes. PC1 explains 83% of the variance and PC3 explains 3.1% of the variance. The question is: if these components are driving the variance, what *biological process* are they representing? What do they stand for? Hoffman says PC1 and PC2 differentiate across the species and PC3 gives the variance over time (i.e., the staged development). This makes PC3 really, really interesting. Just a "couple of hundred" of the genes have high loadings on PC3 but these seem to drive the developmental pattern. What the researchers have to do next is match up the patterns of development across the stages for the genes. Elements of this process seem to be handcrafted (part of the reason for the comment about desktop computing).

Once the relevant gene sequences are identified and matched, the second stage in the analysis simulates embryo development with the different ordered sequences. Agent-based Modelling (ABM) is used to do this. Each sequence is modelled 10,000 times (a standard number of runs in that domain) to see what the stable patterns are. For various historical and technical reasons, ABMs are designed to be run on networked workstations. Hence the other reason for using desktop computing. The rules driving the simulation are given by a tailored operationalisation of 'natural selection'. In his account, Hofmann says he has still to do this work. His summary of what it involves is this:

> **Hans Hofmann**: Every individual cell here is an agent ...... There are very simple rules. For example: If the neighbor is so, then so am I. Only there are two, or three or four rules. They are the same as for a biological process. You generate mutations, the mutations generate different phenotypes, and these are affected by the selection acting in the computer. You can define it. On the computer you can play God, if you want to.
> [p. 125]

The important point of the example and why this use of approach of repurposing data bases has implications for the practices of Biology is that the vast majority of the data Hofmann uses is not his own. That is what makes this research strategy different. What makes it sociologically non-significant (not insignificant) is that all the key issues created by its innovativeness are addressed by well understood socially organised processes. Hoffman and his researchers do exactly what we would expect them to do but in ways designed for *this* particular technique in *their* particular setting. Hoffman lists the 'troubles' he has to deal with as follows.

1. **Quality Control:** This is managed by relying on trust and data massaging. Hofmann cannot control the standards by which individual transcripts are produced so he builds his database from the data of colleagues he knows and trusts; people whose work he respects and who have track records for using standard protocols. Nonetheless, there is variation. To ensure operability across the transcripts, the data are tweaked and adjusted to ensure conformance to the requirements of his algorithms. The need for these kinds of routine adjustments is well recognised within the scientific culture to which he belongs.
2. **Standardised Analytic and Simulation Protocols**. The methods of analysis he relies on, PCA and ABM, are well known. The requirements they have and the trade-offs they bring are common knowledge. You don't have to spell them out. This is as true for the standardisation on correlation as a distance measure and the preference for three principal components as it is for the rule set and structure of the simulation 'competition' in the ABM analysis.
3. **Effacement of these choices**: In the presentation of the analytic results, the 'natural history' of the research is rendered using an appropriately chosen standardised format. 'Posters' are a particular 'rhetorical structure' for a certain kind of scientific presentation and Poster Sessions are routinely organised at scientific conferences.

Whilst the ways Hoffman and his team shape their responses might be distinctive (we have no data on this), the general structures of the actions taken are readily recognisable as typical rational patterns of action, be they in science or anywhere else. They consist in a local variant of what in *Travelling Representations* we called independent, intersubjective validation.

It seems a point made by Gilbert Ryle many years ago applies here. No new 'digital genomic world' is being constructed by these experiments in any sense other than a very attenuated and metaphorical one. What is going on is socially organised practical action taking place in the everyday world of innovating science. This routine practical action utilises culturally given forms. Nothing is hidden and the objects of investigation are not shapeshifters. The modalities of the various renderings deployed in the management of the research are visible to but not talked about by those who understand what is going on and can construe the resulting representations. These modalities are seen but never attended to *unless* they need to be.

Section 4. Realism and The Transmogrification of Fish

The conclusions we have just arrived at underlie why digitalisation is not that interesting philosophically or sociologically even though it might look as if it is or should be. We don't want to 'analyse' the following contributions but rather drop in them as illustrations of the kinds of things we have in mind. They are taken from the discussion of data but could just as easily have been drawn from elsewhere.

> **Gabriele Grameisberger** What you two are describing is a fourfold data concept. There are raw data, metadata, and data about the analysis method, and then the data about all of these data. The layers just keep piling up. When you ultimately speak of the datum you have to unravel all of this, because otherwise you have no chance of understanding this datum. (p. 40)
>
> **Hans-Jorg Rheinberger** What also seems important to me—and this may go back to my own experience as a scientist—is that, in the first instance, the step of data collection is decisive. The data have to come from somewhere. During in vitro experimentation I had the experience that the possibilities for processing at one's disposal are usually a magnitude more precise. If you have to pipette microliters or half-microliters in such an experiment, when you've done that ten times, you have an error margin in the whole process. Afterward you insert the sample in the scintillator, in the machine that counts radioactivity, and this scintillator can measure every sample for you to any place after the decimal. But all of this precision is a waste of time somehow, because the upstream imprecision is much greater. For me, the awareness for what actually hap- pens at the point of collection seems to be a prerequisite for everything else. (p. 51)
>
> **Hannes Rickii** It's a situation that is new for me. I haven't actually done such work before. Therefore, it is like a test, to establish a coherence between the data or the object and its form. But this is preceded by a concept: The point is to rematerialize aspects of this data production. I find all of these abstraction processes—from fish to dataset in the computer—extremely interesting, and I want to keep precisely this chain reversible. For me that is an interesting aspect of biology. One proceeds from a concrete organism, which then dematerializes into scientific work. My question as an artist is: How can I make this material basis that lies behind these many steps, these transformations—the infrastructures, energies, matter, etc.—how can I make it tangible? [p. 53]

The central tenet of realism of any kind is a semantic one. Terms refer. In scientific realism, this is expressed by the claim that in our best attested science, scientific terms refer. The major open issue in the discussion of realism in the applied mathematical sciences is about the realism of the mathematical terms used. These terms (i.e., functions and systems of equations etc.) are mathematical descriptions or renderings of physical objects and processes. They seem to work. We have to take them as real. But we have don't have a good account of how the empirical reality claimed can be provided by the wholly abstract devices used.

Under anybody's theory of science, Behavioural Ecology and Computational Biology are among our best attested sciences. From what our discussants are saying, it seems new representations of biological terms (categories of object and associated activities, for example) are being constructed, collected, stored and subjected to analysis but there is no story for how to track the representational reformulations and preserve the realism of the objects introduced at various points in the research 'pipeline' by using some extended logic of resemblance. When fish activity, say, is recorded as a running stream of video, we are inclined to accept the image in the video is 'more or less' the fish in the aquarium. When the image in the video is processed and stored as a set of temporal co-ordinates and an associated metafile, what has happened to resemblance? And, moreover, is it possible to reproduce (work your way back to) the original fish and their activity from these processed deconstructions? The notion of resemblance and the anchoring of it in what is available under the Natural Attitude and how that is now under strain is the central theme of this particular discussion. It is as if the notions of abstraction and simplification are being defined in terms of decomposition and the realism of the whole conceptual apparatus can only be shored up by re-composable resemblance. What is re-composed must somehow "look like" what you started out with. Interestingly, such re-composing (though perhaps not the criteria of resemblance) seems to concern the video artist as well.

The issue of resemblance is tied to the 'lossiness' of the re-description process and lurking beneath that is the problem of 'literal description'. For logicians, literal description is (formal) description in which terms are fully defined. Mathematical descriptions are the paradigm examples. It is always possible to rewrite a mathematical expression in terms of the definitions of its terms. This rests on the re-write structure of mathematical forms. In Physics the definitions of the basic terms (particle, position, sequence, time etc etc) and many derived terms (force, energy, charge etc) are given mathematically, so it is always possible to walk/work back from a derived description of a phenomenon to the basic terms of the science. The trouble is the re-descriptions used by our biologists don't have this property of two-way constitution (or, at least, they seem not to) and that generates the scientists' discomfort. As the re-descriptions multiply, resemblance and re-composition decline and the route back to the basic depiction/description of the fish in the tank is lost.

Section 5. Re-setting the Discussion.

This problem only arises if one wants to shoehorn non-literal descriptions into procedures designed for literal ones. In her account of applied mathematics, Penelope Maddy [Maddy 2000] makes the point Physics is actually much more relaxed about shoehorning its descriptions in this way than is often admitted or recognised. It happily adopts an approach which is more underdetermined (shall we say more indexical?) and where its basic terms are being drawn from the foundational Kant/Frege (KF) conceptual structure posited for ordinary life. The rigour is not in the literalness of the mathematics but in the systematicity of the analytic procedures (such as enforced strict adherence to the conventional rules by which the mathematics are deployed on the structures of KF categories). This is, of course, very similar to what underpins the efficacy of everyday descriptions. It is not the possibility of reduction to some common definition but conformity to culturally institutionalised (socially organised) precepts which fixes what counts as 'good' descriptions 'in cases such as this'.

If our biologists could give up the shibboleth of the literalness of representations, they might be less discomforted by the character of their disarticulated sequential renderings of their data and as a consequence less open to being seduced by the metaphysical siren song endlessly repeated by Latourian sociologists of science. Re-setting the issues in this way will not solve their practical problems. What it might do, though, is free up some otherwise wasted time and attention which could then be used to address them.

Bibliography

Fischer, P., Gramelsberger, G., Hoffmann, C., Hofmann, H., Rheinberger, H.-J., and Rickli, H. 2020. Natures of Data. Chcago University Press, Chicago.

Garfinkel, H.,. Lynch, M.&. Livingston E. 1981. The Work of a Discovering Science Construed with Materials from the Optically Discovered Pulsar. Philosophy of the Social Sciences, vol 11, No 2, 131–158.

Maddy, P. 2000. Naturalism in Mathematics. OUP, Oxford.