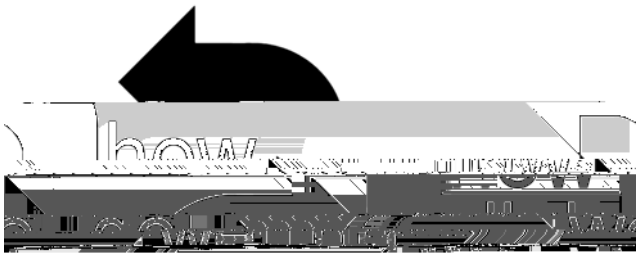
Working Papers on The Nature of Evidence:
How Well Do 'Facts' Travel?
No. 07/06

# Wormy Logic: Model Organisms As Case-Based Reasoning



“The Nature of Evidence: How Well Do ‘Facts’ Travel?” is funded by The Leverhulme Trust and the ESRC at the Department of Economic History, London School of Economics.

For further details about this project and additional copies of this, and other papers in the series, go to:

http://www.lse.ac.uk/collection/economichistory/

Series Editor:

Dr. Jonathan Adams
Department of Economic History
London School of Economics
Houghton Street
London, WC2A 2AE

Tel: +44 (0) 20 7955 6727
Fax: +44 (0) 20 7955 7730

in the past few decades. In addition to the mapping and sequencing of the human genome, among key components of the Human Genome Project (HGP) which officially began in 1990 was the mapping and sequencing of the genomes of non-human model organisms, including mice, nematode worms, flies, *E. coli,* and yeast.[3] James Watson has described the idea to include non-human model organisms in the HGP as his most important contribution to the project.[4] Despite this sort of support from early enthusiasts, some of the more contentious issues raised during the preliminary planning stages of the HGP related to the model organism projects, perhaps most importantly whether genetic sequencing was likely to result in knowledge that was relevant for the understanding and treatment of human disease processes, especially given the large amount of DNA without known function which was often derogatorily termed "junk DNA". Research on model organisms was rarely explicitly defended in the context of the project in its earliest days, perhaps in part because of assumptions about public and political perceptions and lack of ability (or desire) to understand this research, despite their explicit inclusion.[5] These organisms were used in the HGP as a means for developing the various mapping and sequencing technologies needed to study the more complex human genome, thus allowing these technologies to be tested and refined in a simpler, more efficient, and (purportedly) less expensive manner.[6]

---

[3.] See R. A. Ankeny, "Model Organisms as Models: Understanding the 'Lingua Franca' of the Human Genome Project," *Philosophy of Science, 68* (2001), S251-S261.
[4.] R. Lewin, "The Worm at the Heart of the Genome Project," *New Scientist, 127, 1731*

But the genomes of these model organisms also were mapped and sequenced because they were expected to provide a basis for understanding normal gene regulation and human genetic disease, and more generally fundamental developmental, physiological, and other biological processes. Such expectations were based on the idea that many genetic and biological similarities exist between those organisms selected to serve as model organisms and humans; therefore model organisms would provide information that could aid in the interpretation of human genomic sequences and their products. This concept is rooted in the idea that there is conservation of many mechanisms and processes:

> Because all organisms are related through a common evolutionary tree, the study of one organism can provide valuable information about others. Much of the power of molecular genetics arises from the ability to isolate and understand genes from one species based on knowledge about related genes in another species. Comparisons between genomes that are distantly related provide insight into the universality of biologic mechanisms and identify experimental models for studying complex processes.[7]

Both the prevalence and centrality of model organisms in contemporary biomedical research, and claims about their use as the basis for deriving insight into certain common or even universal biological mechanisms, generate an ideal laboratory for examination of epistemic issues related to use of such organisms. In addition, the growing literature within the history and philosophy of science on conceptual issues associated with modelling and representation in science[8] and on

---

their organisms of choice in their own right, which in turn created various epistemic and pragmatic tensions within many laboratories and research programs, a point which I cannot examine in any detail here.

7. F. S. Collins, et al., "New Goals for the U.S. Human Genome Project: 1998-2003," *Science 282* (1998), 682-689, on 686-687.

8. For instance see M. S. Morgan and M. Morrison (eds), (1999) *Models as Mediators* (Cambridge, Cambridge University Press, 1999); S. de Chadarevian and N. Hopwood (eds), *Models: The Third Dimension of Science* (Stanford, Stanford University Press, 2004).

va

surrounding the organism itself.[11] This essay explores some of the techniques and methods used to establish and refine model organisms, but primarily from the point of view, as it were, of the model organisms themselves.

**Background: The Worm**

*C. elegans* is a free-living nematode, around a millimetre in length, with extremely simple behaviours and structures, and a relatively recent history as a model organism.[12] As noted in the Nobel Prize for Physiology or Medicine presentation speech for 2002 which celebrated three worm workers and the "joy of worms", part of what makes it a good candidate for a model organism is that *C. elegans* is "loaded with features".[13] There are two sexual forms, a self-fertilizing hermaphrodite

reproductive system. The organism is transparent throughout its life cycle, making observation of many biological processes possible by various forms of microscopy. The genome of *C. elegans* is approximately 100,000,000 base pairs, which is one-thirtieth the size of the human and twenty times that of *E. coli,* and was virtually completely sequenced as of December 1998.[14]

The choice of *C. elegans* by Sydney Brenner in the mid-1960s and the original pursuit of research focused on this organism primarily at a single institution (the Laboratory of Molecular Biology in Cambridge, England) to which most current-day researchers can trace their own lineages has resulted in a relatively cohesive community often celebrated as a model of scientific cooperation and shared understanding of fundamental concepts.[15] Hence, an analysis of how "the worm" (as it is called by researchers in this area al pursuit o1.5any in the broader scientific community) functions as a model or sm can be used as th s for understanding the epistemic structure underlying most ongoing research in this area.

General examination of the history of organism choice reveals that prospective model organisms typically are selected o1.const ructed

even typicality of their biological characteristics o1.proc esses (though it is hoped that 5any featur es will.prove to be shared or common to other organisms), but primarily due to perceived experimental 5anipulability ao1.tractability. For example, *C. elegans* was chosen specifically for its

---

14. *C. elegans* Sequencing Consortium, "Genome Sequence of the Nematode *C. elegans:* A Platform for Invesl0 13.024pTj1Bi

biological characteristics (among many others) of *C. elegans,* even in comparison to other closely related organisms. In short, the general aim of the original research project was to achieve an understanding of developmental processes in metazoans (animals with bodies composed of differentiated cells, as opposed to protozoa or unicellular animals), and in particular, the development of the nervous system, since it was thought to be the most complex and interconnected system in these organisms.[16]

Brenner wanted to do research with an organism which was experimentally straightforward to manipulate and had relatively basic behaviours and structures, but was not so simple as to be "unrepresentative". The goal was to "optimize" an organism, in large part through making a careful organismal choice to start, rather than focusing on achieving standardization once in the laboratory via inbreeding and other typical techniques. Brenner and most subsequent worm workers in the early years of the research implicitly assumed that although *C. elegans* is simple, it is similar to all (or most) of the more complex members of the metazoa in terms of the genetic control of cellular differentiation. In particular, the genetic control of the development of the structure of the nervous system was thought to be likely to have shared fundamental mechanisms, in large part because of an implicit assumption of genetic conservation, particularly of essential processes.

be compared, in order to articulate variations and differences in various features. The use of this form of reasoning in is perhaps most familiar from basic genetics: the first step in the underlying strategy is to select and establish a "wild type" for the organism (taken as a standard from among other possible wild types available in nature) against which other genetic variants or abnormal types can be compared. Despite its name, the wild type may not be the most common, frequent, or even a "normal" version of the organism; sometimes it is simply the first strain that was discovered on which subsequent research has been based, but is oftentimes the easiest to manipulate experimentally. These experimental organisms of course are "natural", inasmuch as they are still actual, living, concrete organisms, and have been "selected from nature's very own workshop".[18] However, the carefully selected wild type is, in this sense, an idealized model of actual organisms in nature, since oftentimes they end up differing considerably from those highly rarefied beasts that remain isolated in the laboratory, particularly as a model organism comes to be more widely used.[19] Thus modelling occurs in most obviously in the establishment of the wild type, which is an essential first step to establishing and using something on an ongoing basis as a model organism. Without this, it is not possible to have a "norm" against which "abnormal" (or more precisely, that which is variant) can be compared, in terms of genetics, developmental lineages, and so on. So a worm that is abnormal in movement might be detected by comparison of the paths that it traces in response to a stimulus to those traced by a worm held to be "normal".

---

[18.] E. F. Keller, *Making Sense of Life: Explaining Biological Development with Models, Metaphors, and Machines* (Cambridge, Harvard University Press, 2002), p. 51.

[19.] In organisms where there is ongoing flow over time between the laboratory and the field or the wild, the amount of idealization in the model may be reduced, or more precisely, there may be more than one strain or variant that is held as a norm; however particularly with genetic model organisms (those selected primarily because of their power for genetic analysis, which is my focus in this essay), it is essential to settle on (and persist in using) one wild type.

A second way in which modelling occurs is in the establishment and use of what I have called elsewhere a "descriptive model".[20] The term "descriptive" is utilized to capture the idea that these sorts of models are descriptions which serve as prerequisites to explanatory questions; their articulation often is not motivated (at least immediately) by their future potential explanatory value. Thus in model organism work, there typically is an extensive research phase in which a descriptive model of the organism is developed. Consider, for example, the articulation of the "wiring diagram" of the neural connections within *C. elegans.* This model was a paper (and later computerized) series of drawings, which resemble electric circuitry diagrams.[21] The overall diagram was constructed by combining wiring diagrams from several individual wild type worms, not only because of practical or experimental limitations, but because it was deemed necessary to eliminate what seemed to be individual neural differences (even between genetically-identical organisms) in favour of a canonical nervous system. The wiring diagram is based on an abstract model of the worm in terms of the typical or usual neural connections exhibited not by any one specimen alone, or by numerous individual organisms, but by a more abstract construct hybridized from a few individual specimens. The wiring diagram thus is a model of the worm in terms of the typical or usual neural connections exhibited not by any one specimen taken by itself but by a very precisely derived type of construct.[22] This descriptive model is compared to the wiring diagr

for worms that are variant or abnormal in neural patterns in order to assess possible connections between variations in genetic sequence and in neural structure, and eventually to

researchers, among other purposes.  To begin, it is helpful to provide a brief overview of the general form of case-based reasoning as used in

variants or errors in what was assumed to be the shared or common attributes (genetic, physiological, and otherwise) among healthy individuals are discovered. Thus the index case of the normal and of the disease condition often are constructed (and re-constructed) in terms of each other as more knowledge is gathered. What is essential in this form of reasoning is the feedback loop that exists between the descriptive model of the normal and the descriptive model of the abnormal condition. Newly-acquired evidence can change what is considered to be the index case or whether something should be considered to be a unique case at all.

Thus these cases are models inasmuch as, although they originate from some actual observed instance in the first place, once they begin to be disseminated and used, they become idealized away from particular details of the observed phenomena. They serve as intermediaries 101 Tw 13.02 0 0 13.0

model, and yet patients can still be identified as having a condition or being an instance of that particular disease category or case.

### Model Organism as Cases

The practices of contemporary biological science have (potentially conflicting) goals that are similar to those found in the practice of the medical sciences. There is a desire to get to the fundamental biological characteristics shared by all living things, be they biochemical, genetic, developmental, or neurobiological processes. At the same time, biologists are aware that any model system or organism selected for research may be problematic and atypical, particularly inasmuch as such systems are proving to be complex in ways previously that might not have been anticipated. The previous section on *C. elegans* as a model organism has shown several ways in which the organism as studied by biologists is an idealized entity or a model. The epistemic strategy of using the models as cases allows them to serve as a means of control of complexity, a way to create an appropriately simplistic yet descriptively rich basis for future studies and more traditional hypothesis testing, experimentation, and explanation.

Different aspects of a model organism thus can be viewed as index cases on which comparison to variant and abnormal instances of the same organism. So for instance, the wild type of the “natural” organism serves as an index case, in that it establishes a genotype which comes to be understood as “normal” and serves as the basis for comparison to subsequent cases of abnormal or variant genotypes. Similarly, the wiring diagram captures another sort of basic index case, to which variations in neural structure can be compared. Among the key foundational assumptions used to determine what counts as the relevant or most useful base index case for an organism are the anticipated degree of

practice.[25] Thus there is a creation of an epistemological space or framework within which to ask questions. However, as bluntly stated by a commentator on medical reasoning, “with higher organisms, and especially with patients, it becomes hopeless to attempt to create complete descriptions…This is a kind of epistemologic surrender and consists in simply ignoring many of the things that could be truthfully said in order to say what must be said”.[26] Both in medicine and in biological reasoning from model organisms, complexity, completeness, and perhaps “naturalness” are sa.02305TjETEMC /P <</MCIDcEMC7502 2 0 0 13.02 122.6eTj 295.3201

human genome (or other, “higher”, organisms) and prove fruitful for understanding the functional properties of these sequences.  Finally, the eventual goal is to understand the higher level, phenotypic

disanalogies and their import).[28] Much rhetoric surrounding model organism research unconstructively obscures this interplay and hence misrepresents the potential limitations of even good models. In other words, providing a model requires an interaction between the model and the object of interest being modelled, or between the base index case and the case of interest, including construction of similarity relations, which are impossible to devise without a detailed description of the process to be modelled (which in this case includes the functional properties of the sequence).

Case-based reasoning is an epistemic process that is far from straightforward and may seem to fail to allow us to obtain the usual results we expect in science, inasmuch as it fails (at least initially) to produce unified theories or mechanistic explanations, but instead results in a form of scientific understanding (perhaps of a weaker sort than our traditional theories and explanations) which is constantly evolving, incomplete

About the Author:

**LONDON SCHOOL OF ECONOMICS**